Omixon Target User Manual

Omixon Target User Manual

Introduction

Introduction

Omixon Target ('the tool') is a suite of software for analysing targeted sequencing data from next generation sequencing (NGS) platforms.

The main vision behind this tool is to help laboratories move towards using NGS data for the analysis of diagnostic targets.

There is a Quick Start Guide for Omixon Target available on the Omixon website (http://www.omixon.com/omixon-target-quick-start-guide/).

Omixon Target is for Research Use Only. Not for use in diagnostic procedures.

Uses

Analysis of Diagnostic Targets

HLA Typing

Omixon Target has a simple HLA Typer for NGS data. This tool is easy to set up and run, and multiple samples can be analysed together. It provides both a visual summary, and tables of reports with statistical confidence measures for the accuracy of the results.

Data Analysis

Omixon Target includes preconfigured support for analysing some of the most common human diagnostic targets such as CFTR, and BRCA, however you can also configure the tool to analyse any gene or region of interest. The underlying mapping, alignment and variant calling algorithms are intended to meet the high precision and analysis quality control requirements of diagnostics labs. You can easily move from a high-level mutation summary table to inspect the underlying short read data, and based on this visual inspection you can manually approve variants. The tool helps to identify amplification artefacts and other error sources using a simulation-based validation track as an analysis control. The tool can be used for multi-sample or family trio comparative studies.

Discovery

Omixon also use the tool to support collaborative genomic biomarker discovery projects.

Sequencing Technologies

Omixon Target supports sequencing data from these major sequencing technologies:

- Illumina
- Ion Torrent
- Roche 454

Modules

Omixon Target has a modular structure

Omixon Target consists of a free Genome Browser Module, with some optional Modules.

This means that Omixon Target can simply be used as a Genome Browser, if desired (i.e. for visualization of NGS data analysis results, in SAM/BAM and VCF/GFF/BED formats).

Free Sandbox Genome Browser Module

This is always included as standard with Omixon Target. It includes some data management functions, as well as a full-featured NGS Genome Browser. This module is available free-of-charge.

The Sandbox Genome Browser Module works on Windows, Linux and Mac OS X.

Standard Data Analysis Module (Optional)

This module includes a range of general genomic data analysis algorithms, including Omixon's own aligner and a full GATK variant call pipeline. It also features variant effect prediction, an approval work flow (for automated or manual filtering and approval of variants) and an automated coverage analysis and coverage annotation pipeline.

The Standard Module works on Windows (64 bit recommended), Linux (64 bit recommended) and Mac OS X.

Alignment

There is one short read aligner available, Omixon's own ORM (Omixon Read Mapper).

Variant Calling

The variant calling is provided by a GATK pipeline (using the Broad Institute GATK variant caller). This gives excellent results with Illumina data.

Variant Effect Prediction

There is the option to import an SnpEff database for variant effect prediction. This will be done automatically for HG19 if any of the built-in profiles (BRCA, CFTR) are used, or if any profiles are created for HG19.

Pro Data Analysis Module (Optional)

This includes all the features of the Standard Data Analysis Module, including all the standard genomics data analysis algorithms and the approval work flow. This Module is an alternative to the Standard Data Analysis Module, for Linux users.

In addition, it also includes integrated access to a number of third party tools, including the BWA aligner, a better variant call pipeline for Ion Torrent and Roche 454 data (based on the SamTools mpileup and some supporting scripts), a demultiplexing import and an sff import feature.

The Pro Module is only available for a 64 bit Linux operating system and requires some manual installation and configuration of third party tools in addition to the usual installation steps.

Demultiplexing

There are some extra tools for automated demultiplexing of Illumina and Roche 454 data.

Alignment

There is an extra short read aligner, BWA.

Variant Calling

There is a another variant calling pipeline based on SAMtools mpileup, which gives better results for Ion Torrent or Roche 454 data

HLA Typing Module (Optional)

Omixon Target contains simple, fully automated HLA Typing algorithms which have been developed to work for reads produced by all three major sequencing platforms. It is possible to do HLA typing from any kind of experiment, including targeted sequencing, whole exome or whole genome experiments.

The HLA Typing Module works on Windows (64 bit recommended), Linux (64 bit recommended) and Mac OS X.

This Module contains only the HLA Typing feature of Omixon Target (i.e. it has none of the data analysis features).

The HLA edition has no up-front license fee. The HLA Typing Module offers two licensing schemes:

1. A credit-based pricing scheme with no up-front license fee, with limited use (one credit is consumed per sample analysed, more credits can be purchased on demand). 40 free HLA typing credits are included in the evaluation version.
2. An annual fee based license, which allows unlimited HLA Typing. Please contact sales@omixon.com for a quote.

Adding and removing Modules

Note that it is possible to add and remove Modules to and from your Omixon Target installation. The Modules are controlled by a license file, so if you would like to alter your installation then you should contact Omixon and we can generate a new license file for you, which can be imported within the Settings function in the tool.

Key Concepts

The section covers the main concepts for the **Data Analysis** modules.

Target

Data Analysis is built around the concept of an analysis target. This can currently be one or more genes, exons or amplicons.

There is a Target Dashboard to manage the targets within the tool. The annotations for a Target can be imported directly from a bed or gff file.

While using a target is recommended it is not actually mandatory within the tool.

Hierarchy of Experiment, Analysis and Sample

In order to help with organising the data belonging to the samples that are sequenced, there is a simple data management hierarchy in the tool.

The sequencing data will belong to a 'Sample'. A number of samples can be grouped together within a single 'Analysis'. And finally, a number of analyses can be grouped together within a single 'Experiment'.

The configuration of shared data, such as:

- the target regions for the investigation
- which reference species genomes will be used (and the imported reference sequence data and known variations)
- and the sequencer used for the sequencing of the data

can be done at both Experiment and Analysis level.

If desired, a more generic configuration can be created within an Experiment, with a more specific configuration chosen within each Analysis. For example, it might be desirable to separate the Analyses by sequencer, if multiple sequencers are being used. It's also possible to use a sub-set of targets within the Analysis, for example the HLA-A gene could be the target of one Analysis, and the HLA-B gene could be the target of another.

There is an Experiment Dashboard, an Analysis Dashboard and a Sample Dashboard to manage each of these.

Profiles

The easiest way to use Omixon Target for data analysis is to use one of the pre-built, pre-configured Profiles. This already contains a Target definition and all the reference data required for a particular Target. Adding a new profile will cause the data to be downloaded and automatically configured with the tool. It will also create and configure a starting Experiment, Analysis and Sample.

Reference Genomes

One or more reference genomes can be added to each Experiment. One or more 'contigs' (usually chromosomes) can be chosen for each reference genome. The reference genomes and/or contigs can split across the Analyses.

The sequence data for the reference genome (fasta format) will need to be imported into the tool. For the best variant analysis results, the dbSNP (known mutation) data for the reference (if available) can also be imported into the tool.

There is a Reference Genome Dashboard to manage the data for the reference genomes.

Samples

Each 'Sample' will generally consist of the sequencing data or results for a single individual. There are three possibilities for importing sample data:

- The 'raw' sequencing data (short reads) can be imported into the tool (in fastq or sff format), after which it can be mapped and aligned against one of the reference genomes (one or more contigs) and variants can be 'called' (detected) - all in a single step.
- It is also possible to import short reads that have been mapped by another tool and just do the variant detection - or simply just browse the data in the Genome Browser.
- Finally, variant call results can be imported alone, without any short read data at all, and visualized and analysed within the tool.

Once variants have been called, there are a number of downstream analysis steps that can be performed, including browsing the results in the Genome Browser, analysing the results within the Analyse Sample Dashboard, 'approval' of variants via the Approval Dashboard, and creating and viewing reports in the Report Dashboard.

Mapping, Alignment, Variant Call

These functions are only available within one of the optional Data Analysis Modules (Standard or Pro).

One of the primary goals of the tool is to identify variants within the sample.

The tool includes sequencer-specific algorithms for mapping and aligning NGS short reads against a reference sequence, and then calling variants against the aligned short reads using a GATK pipeline (following the recommended best practises of the Broad Institute).

The mapping and alignment algorithms used are currently Omixon's own. There are separate algorithms for Illumina, 454 and Ion Torrent data, which include different error models for the different sequencers. One future plan is to run not just Omixon's algorithms, but some other alignment algorithms as well, and to compare and/or merge the results of this double alignment.

It is also possible to simply import already mapped short read data and just run a variant call, or even just import already called variants into the tool, for downstream analysis.

Genome Browser

There is a genome browser included in the tool. This allows you to browse the selected target regions, manually inspect the variants and short reads, see a consensus sequence, and see the coverage of the short reads. Multiple samples can be browsed together.

Analyse Sample

It is also possible to import 'expected' variants into the tool. There is a function called 'Analyse Sample' which will automatically compare the expected variants with the actual variants found (either imported or via the map/align/call variants functions). These expected variants can come from previous analyses (for example with Sanger data), or they are also useful if the data being analysed is simulated data, with known results.

The Analyse Sample function also generates a target based coverage report, containing some annotations (e.g. for zero coverage regions within the defined target) and basic coverage statistics.

Sample Difference

If multiple samples are selected in the Analysis Dashboard they can be compared against each other using the 'Compare Samples' function, which opens the Sample Difference screen. On this screen, variants for all samples are shown within a single table. The selected samples can also be browsed together and compared visually in the Genome Browser.

Manual Approval

This function is only available within one of the optional Data Analysis Modules (Standard or Pro).

It's possible to 'approve' or 'reject' variations within a Sample. Approved variations will become candidates for downstream analysis, such as HLA typing. Rejected variants will be ignored by downstream tools. Approval and rejection can be done either by the Automatic Approval or Automatic Rejection functions, or via a manual approval process within the Approval Dashboard screen.

Bioinformatics Notes

Reference Sequences

By default, Omixon Target is set up to use HG19 as the main human reference sequence. It is however possible to configure the tool to use any other reference sequence.

Alignment Tools

These tools are only available in the optional Data Analysis Modules (Standard and Pro).

The main underlying tool for the 'Map and Align' step is the Omixon Variant Toolkit. This is also available as a standalone, command line tool. The Toolkit uses a 'properties' file for it's parameters, and the same file can be used to transfer 'advanced' parameters to the underlying Toolkit while running the 'Map and Align' step within Omixon Target. The Toolkit has it's own readme file, which is also available for download via the toolkit page on the Omixon web site: https://www.omixon.com/omixon/abouttoolkit.htm

BWA is now available in the Pro Module.

Variant Calling Tools

These tools are only available in the optional Data Analysis Modules (Standard and Pro).

Omixon currently uses open source, third party variant calling tools. The only one included in the Standard Module is the GATK variant caller from the Broad Institute. Also included in the Pro Module is the SAMtools variant caller (mpileup), which gives better results for Ion Torrent and 454 data.

System requirements

Omixon Target is supported on 3 platforms:

- Mac OS X
- Linux
- Windows

The recommended hardware requirements for all the optional modules are the following:

- 64 bit multi-core CPU with 64 bit operating system,
- at least 8GB memory,

- storage space requirements mostly depend on the size of data usually used for analyses.

The memory requirements depend on the size of the reference genomes being used.

The memory limitation of 32 bit operating systems is the primary restriction for not fully supporting these. For the Genome Browser Module only or for demonstration purposes with small genomes, the tool can still be used on a 32 bit operating system with up to 4GB memory.

An Oracle Java 6 Runtime Environment (JRE) is required, this is always included inside the installer.

Installation guide

There are separate installation packages for all three supported operating systems.

The following sections contain the descriptions for the installation steps required for each operating system.

The Omixon Target Pro version of the product has a separate section to summarize installation and setup instructions related to this specific type of distribution.

After installation confiuration steps for the server and client application of Omixon Target Server are also described separately.

Windows installation guide

We provide two versions of the installer package for Windows operating systems, both are bundled with a Java Runtime Environment (JRE):

- 64 bit version, named omixon-target_windows-x64_1_4_0_with_jre.exe;
- 32 bit version, named omixon-target_windows_1_4_0_x86_with_jre.exe.

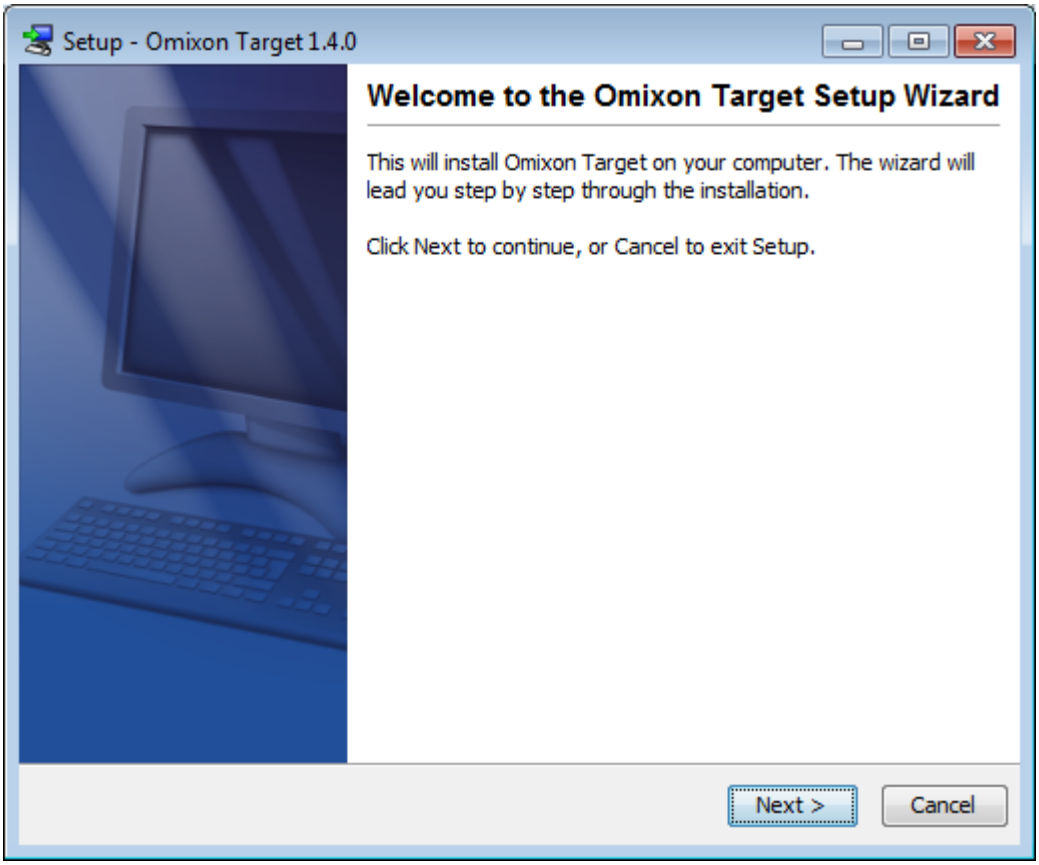
We have tested Omixon Target with Windows 7, Windows XP and Windows 8.

The 32 bit operating systems are not officially supported (not fully tested), but that does not mean that the application cannot run. The current limitation of 32 bit operating systems is mostly related to the amount of memory available.

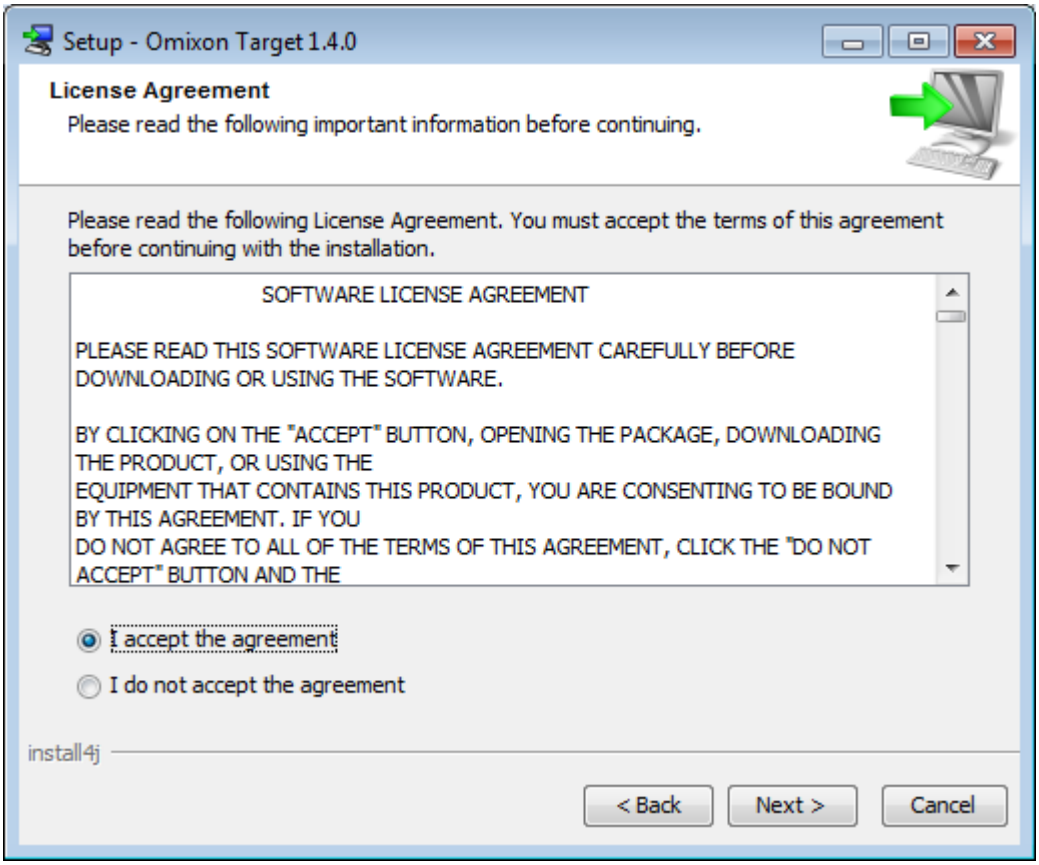
Installation steps

Launch the executable.

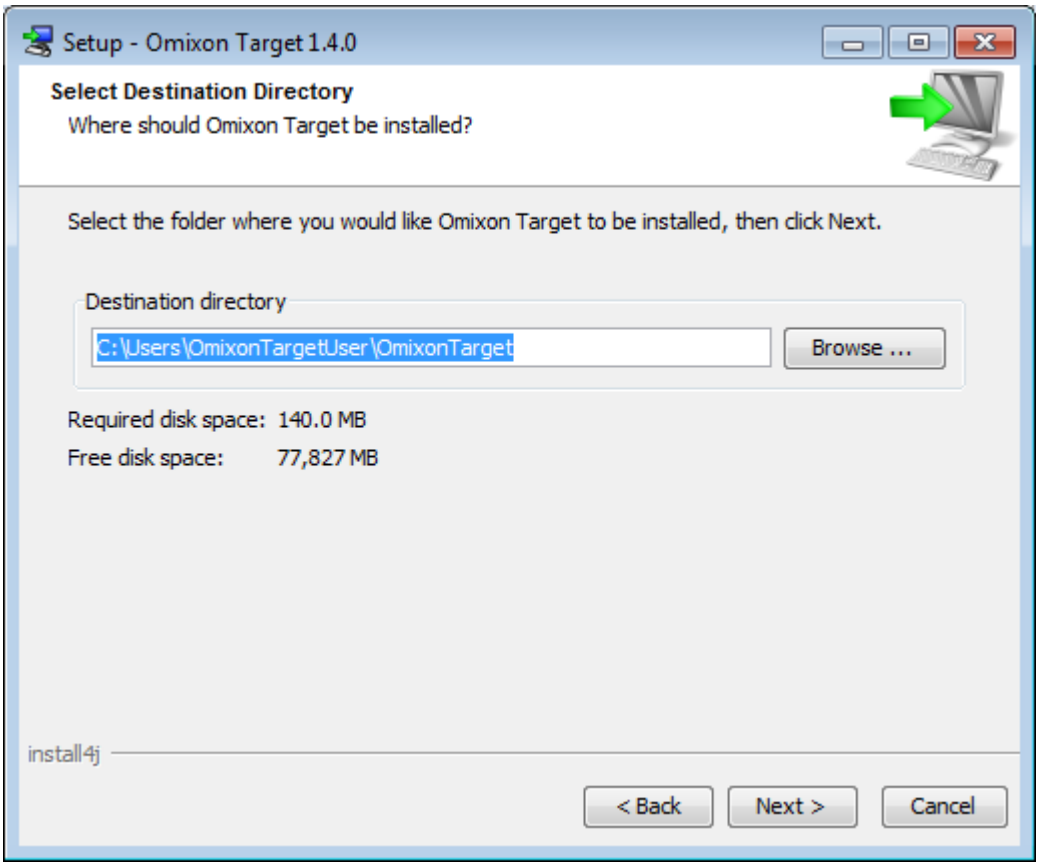
After launching the installer, Welcome dialog of the setup wizard appears.



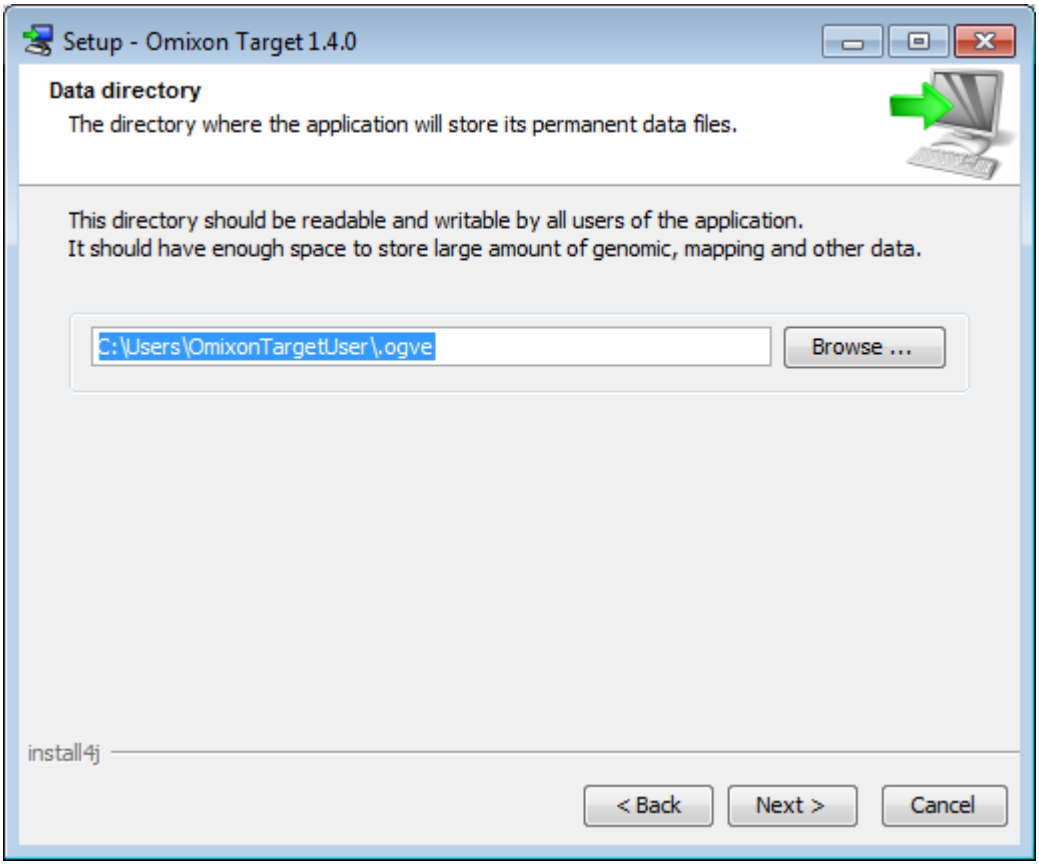
Please, accept the software license agreement to proceed with the installation.



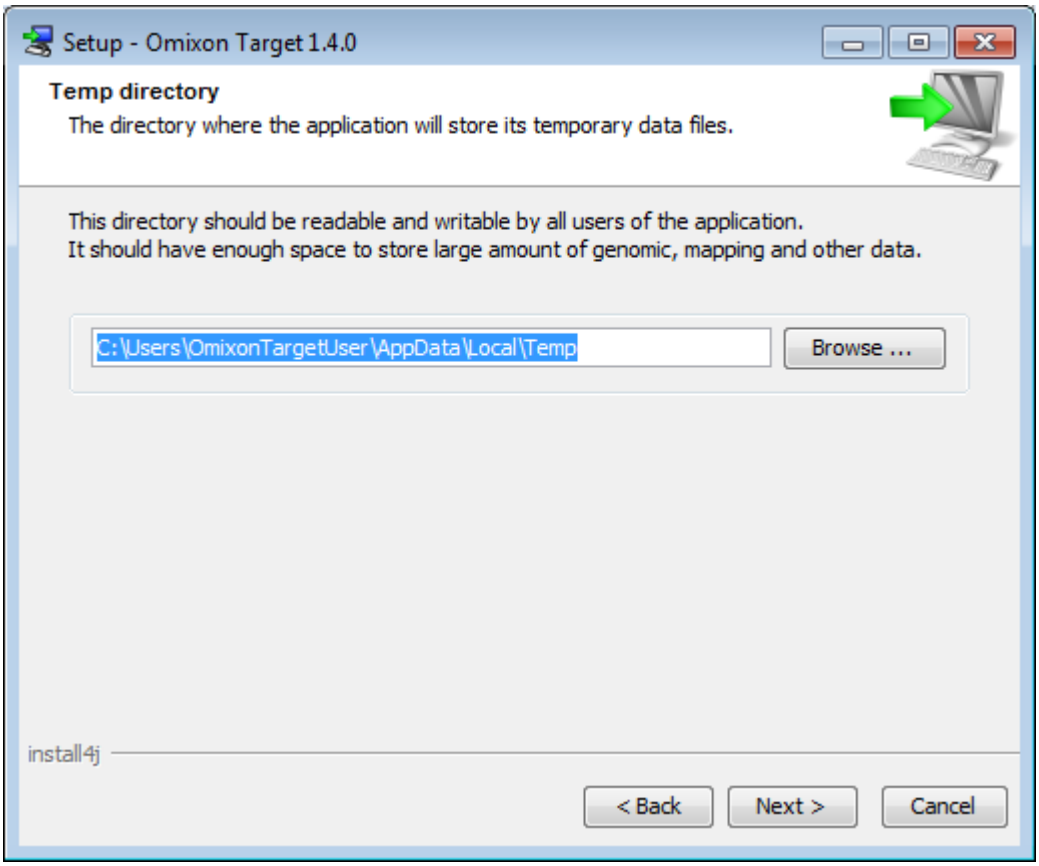
Select the path where you would like the application to be installed.



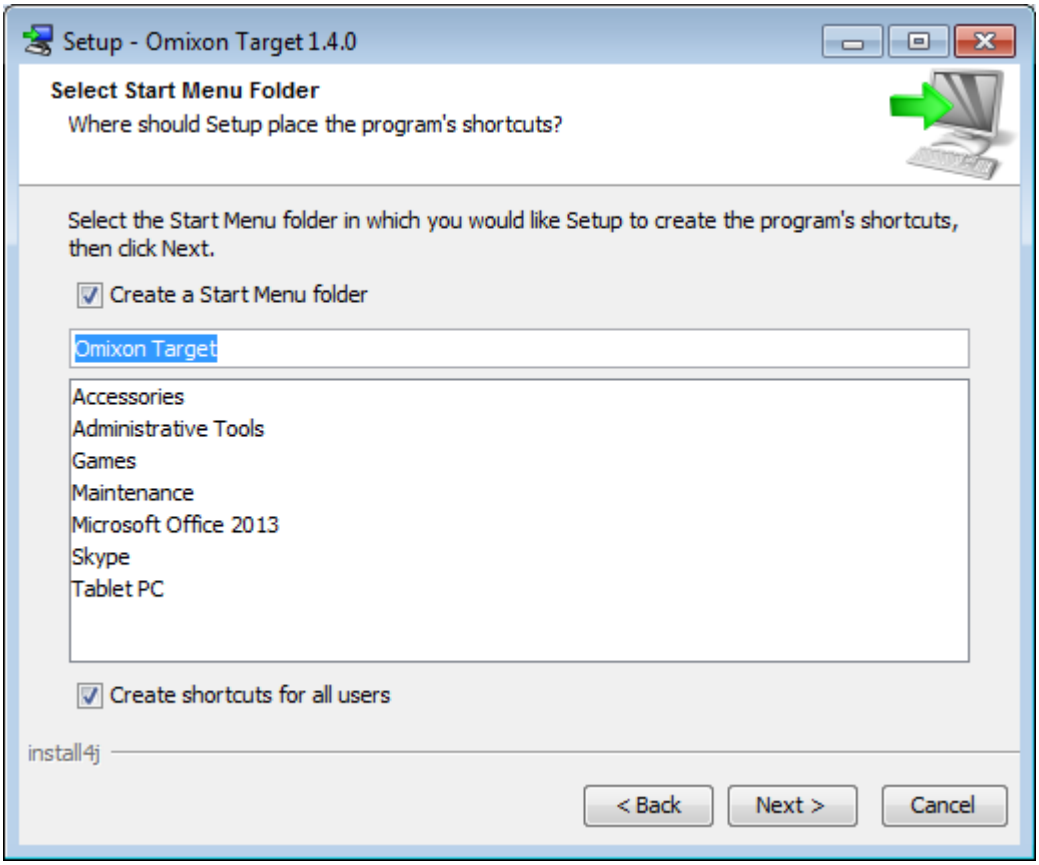
Select data directory where the permanent data files will be stored.



Select temporary files directory.

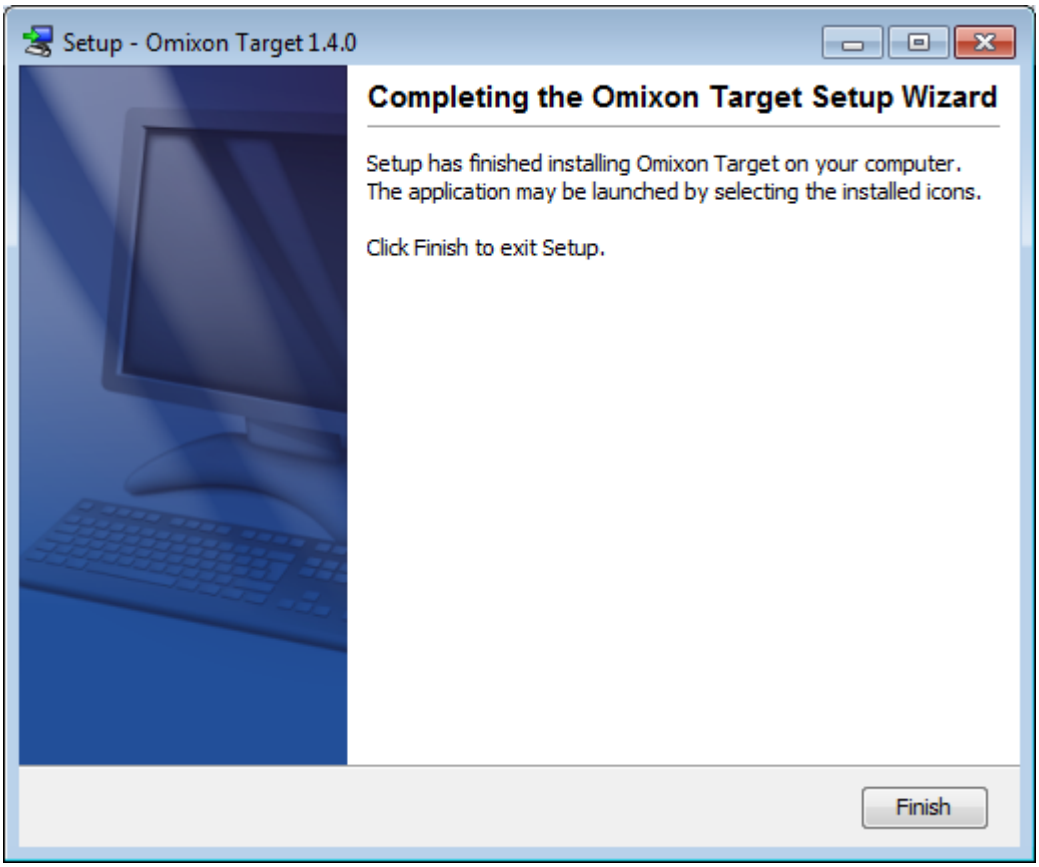


Enter name for Start Menu item. Use the checkboxes to make it visible for all users of the operating system or to not create Start Menu item at all.

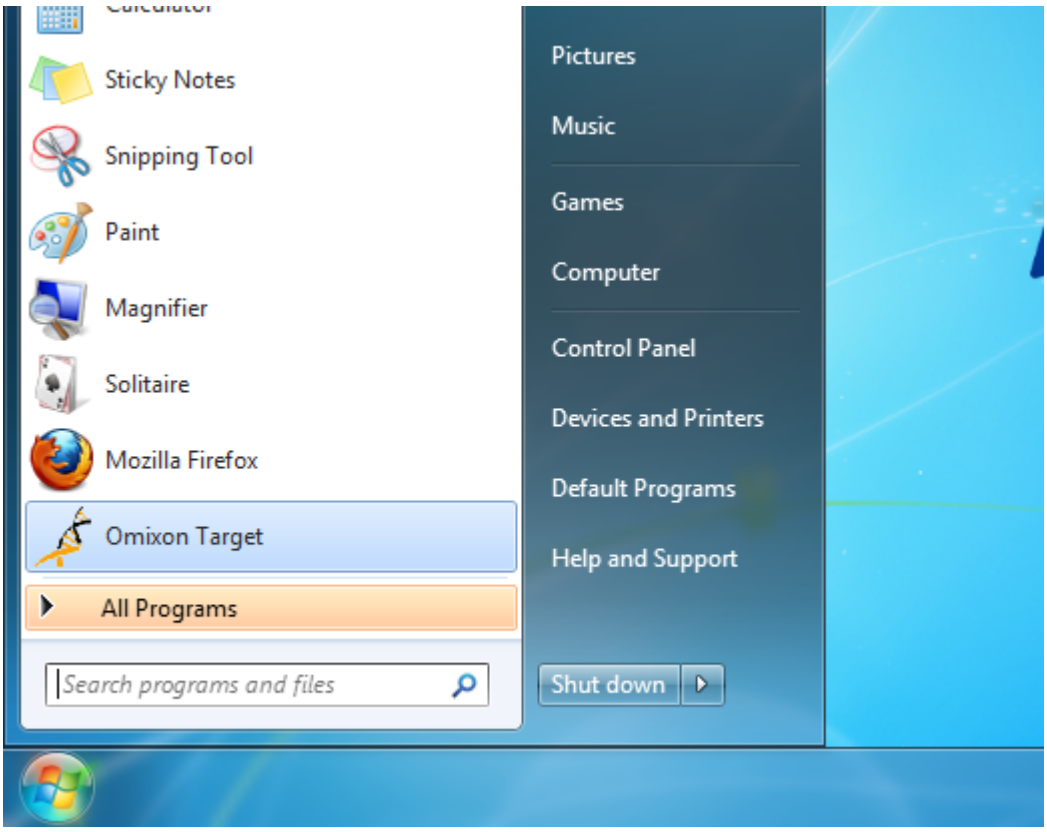


After clicking the Next button, the installation process starts.

When it is finished, the Completing the Omixon Target Setup dialog indicates the success of the installation.



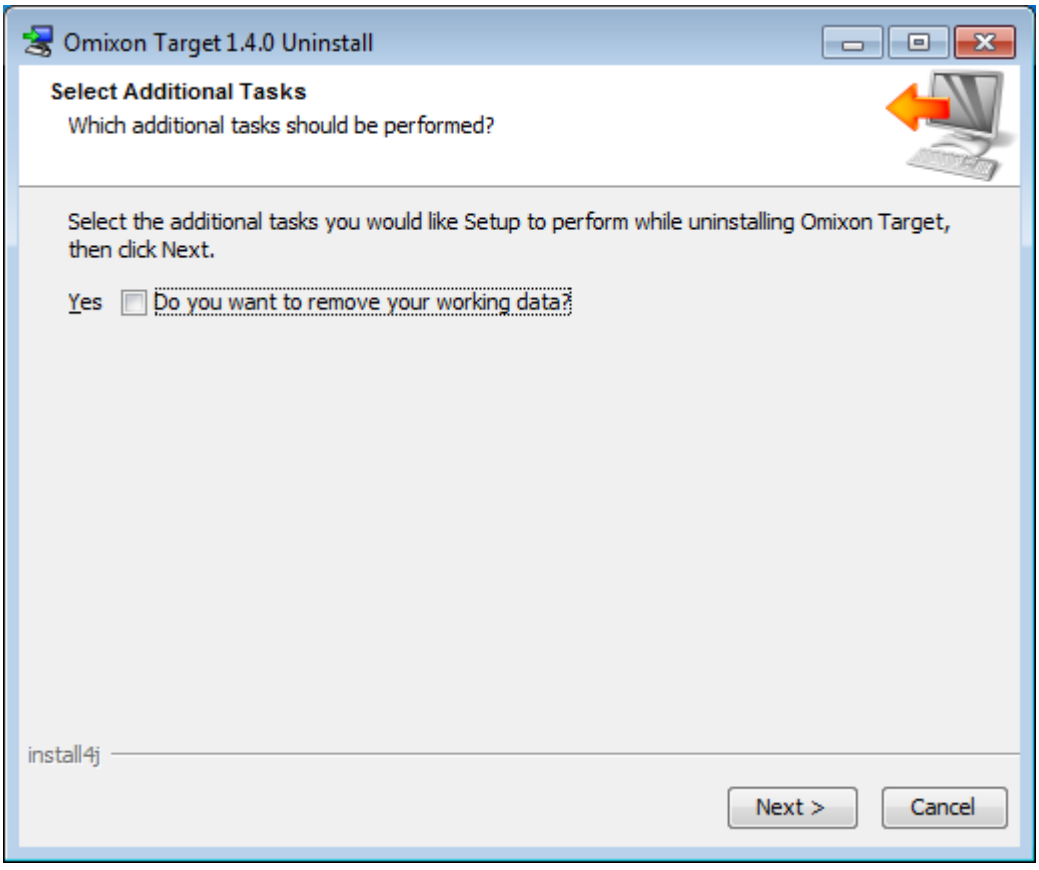
If you've decided to create a start menu folder in the previous step, 'Omixon Target' shortcut is available now. Omixon Target application can be started by clicking on this startup icon or running the executable from application directory.



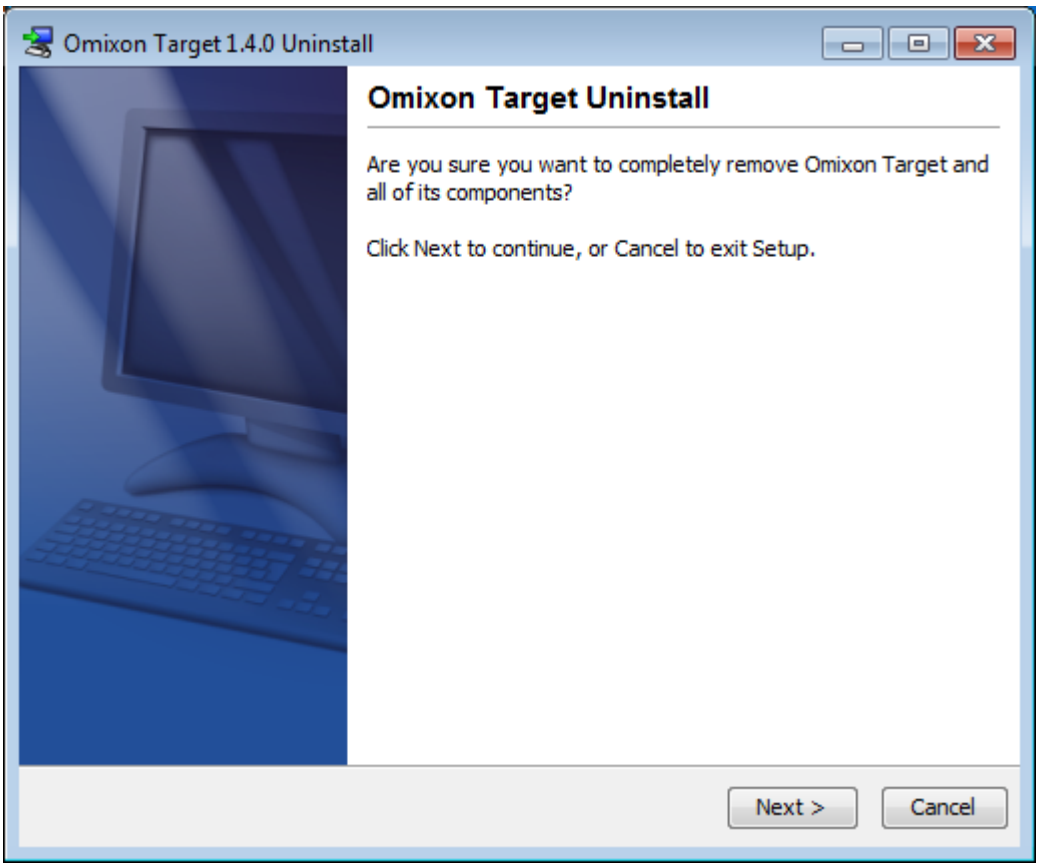
Uninstalling the product

If you want to remove Omixon Target, an uninstaller executable is available in the directory where the application was installed. If you have created Start Menu items during installation, 'Uninstall Omixon Target' icon launches the uninstaller.

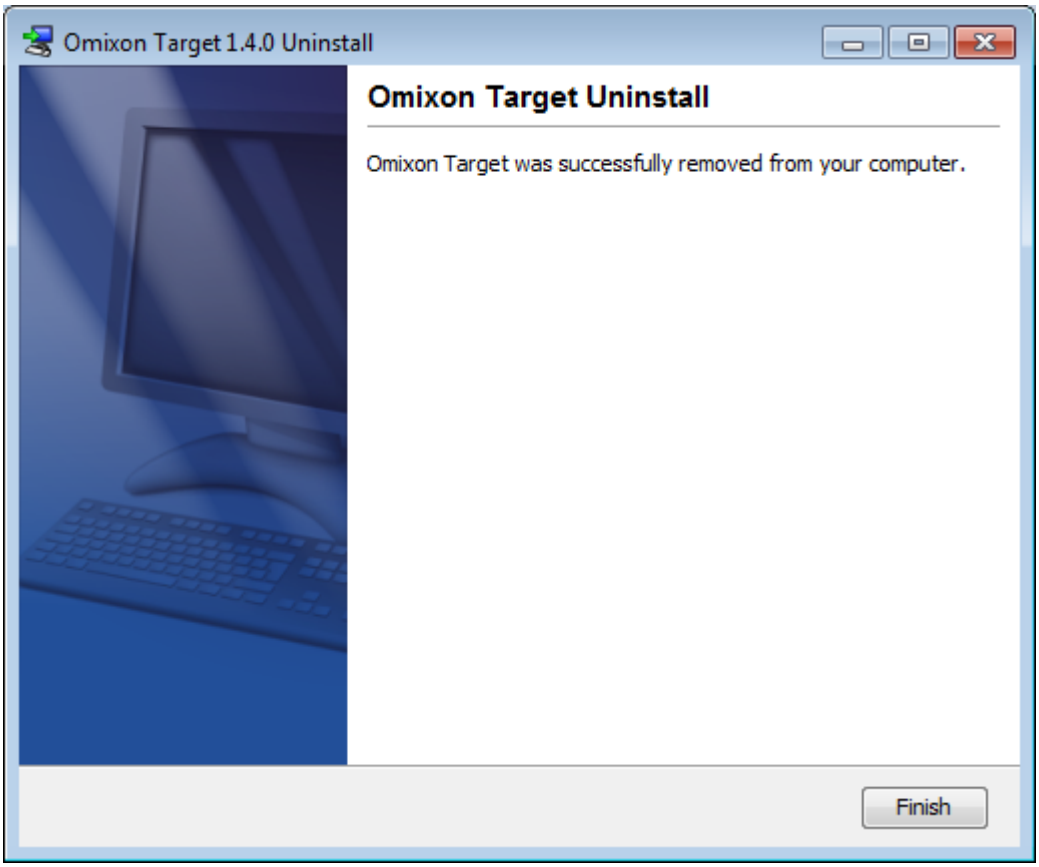
If you intend to use future version of Omixon Target and you don't want your work data to be deleted, keep the checkbox unchecked. If you wish to cleanup the working directory, then selet the option 'Yes'.



Click 'Next' to proceed with the uninstallation process.



Finally, the installer informs you about the outcome of uninstall.



Mac OS X installation guide

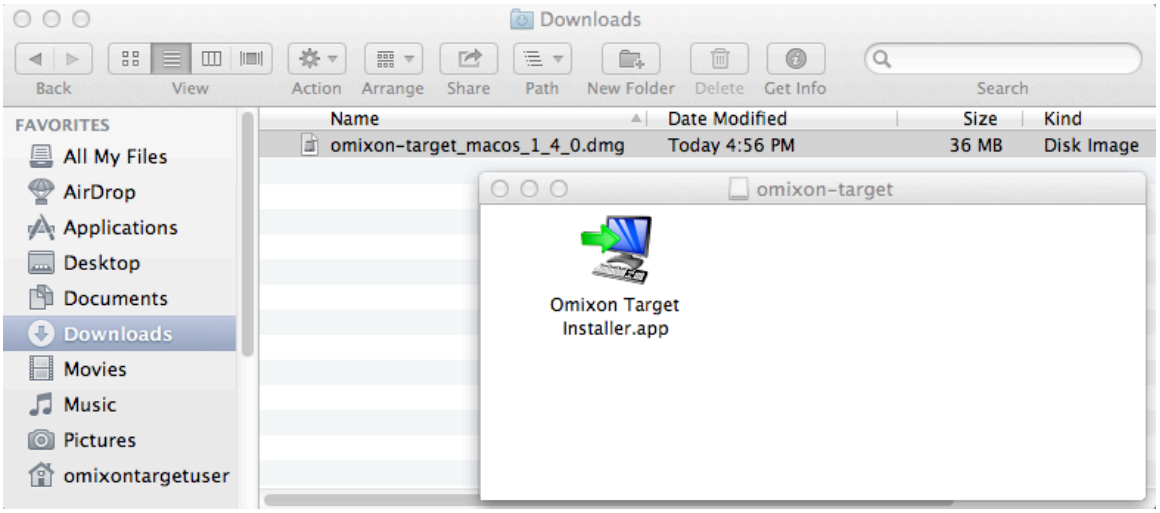
As Mac OS X is an exclusively 64 bit operating system and it always has a proper Java Runtime Environment for our tool, therefore the installer does not contain Java Runtime Environment (JRE).

- omixon-target_macos_1_4_0.dmg

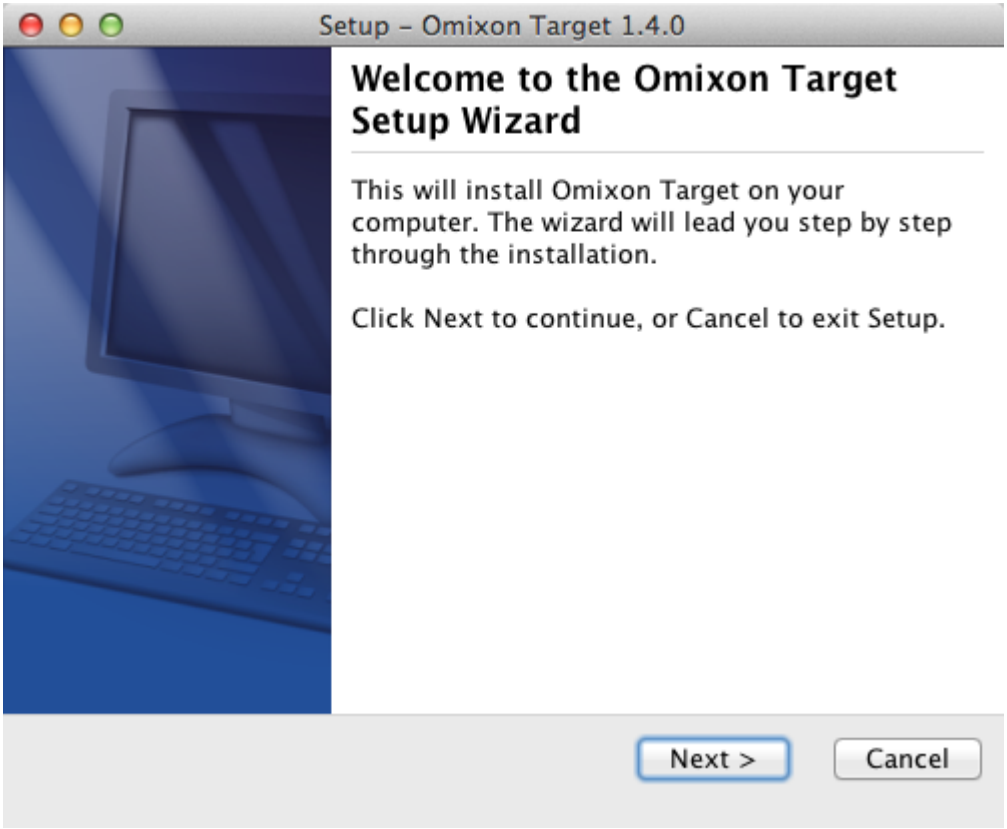
If you don't have Java installed on your Mac, it will be automatically downloaded during installation.

Installation steps

Open the installer which is packed within a DMG archive. Click on the 'Omixon Target Installer.app' icon.

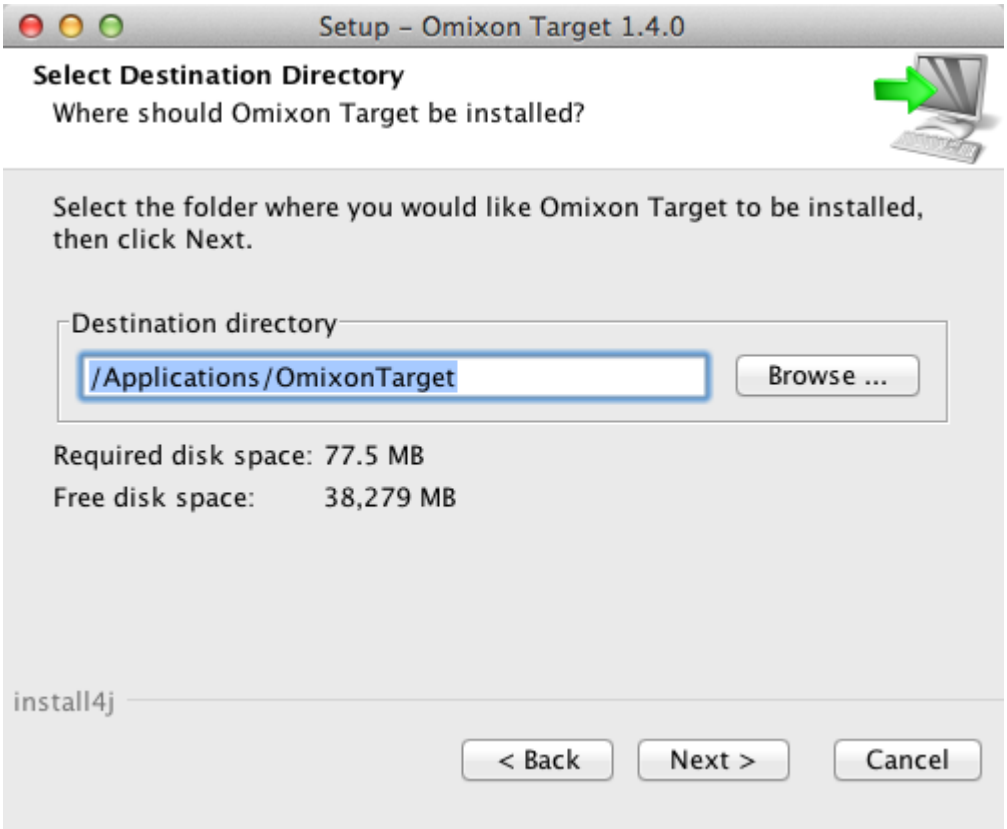


After launching the installer, Welcome dialog of the setup wizard appears.

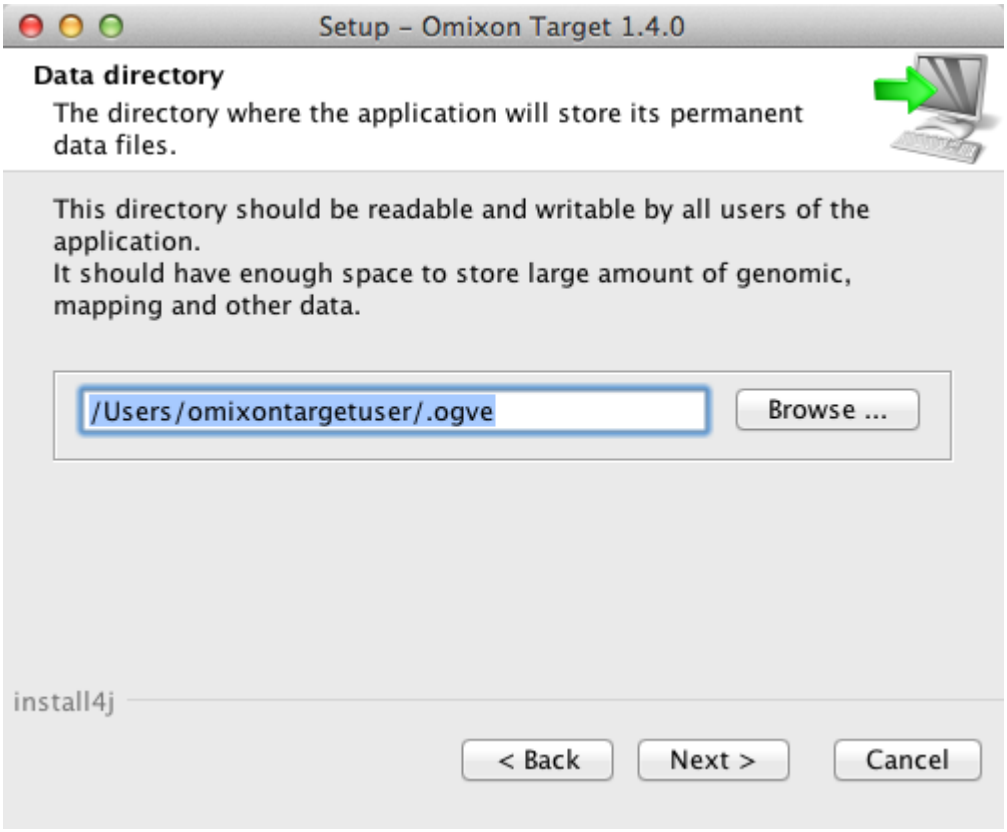


Please, accept the software license agreement to proceed with the installation.

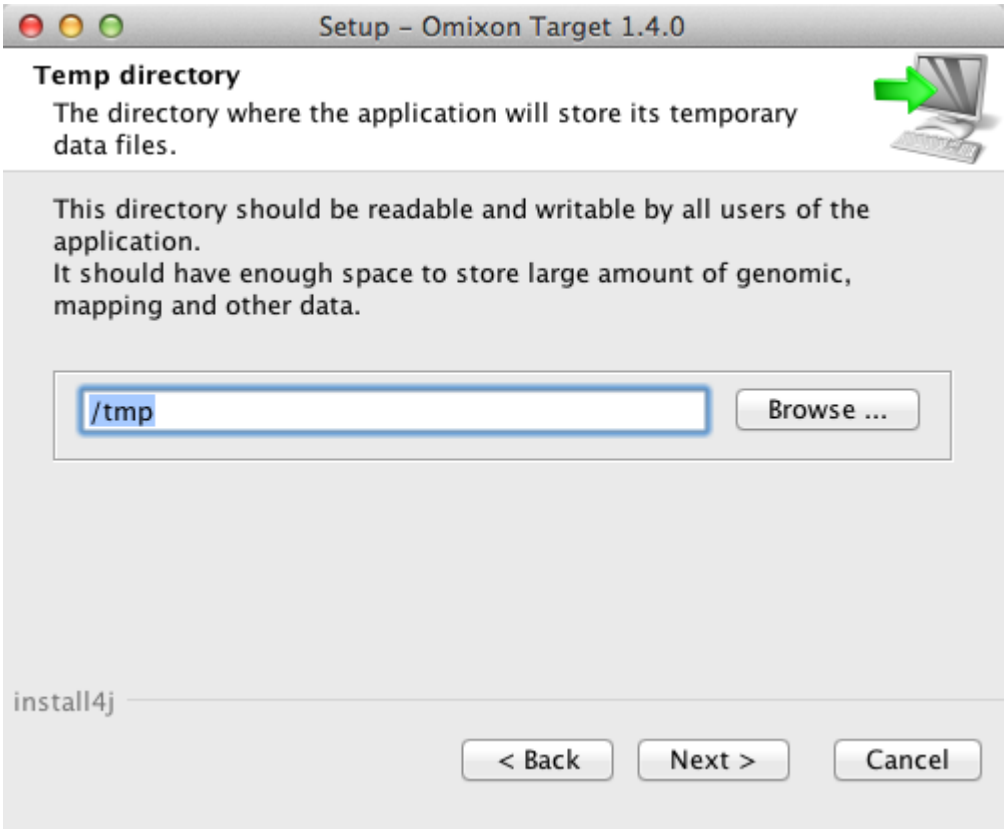
Select the path where you would like the application to be installed.



Select data directory where the permanent data files will be stored.

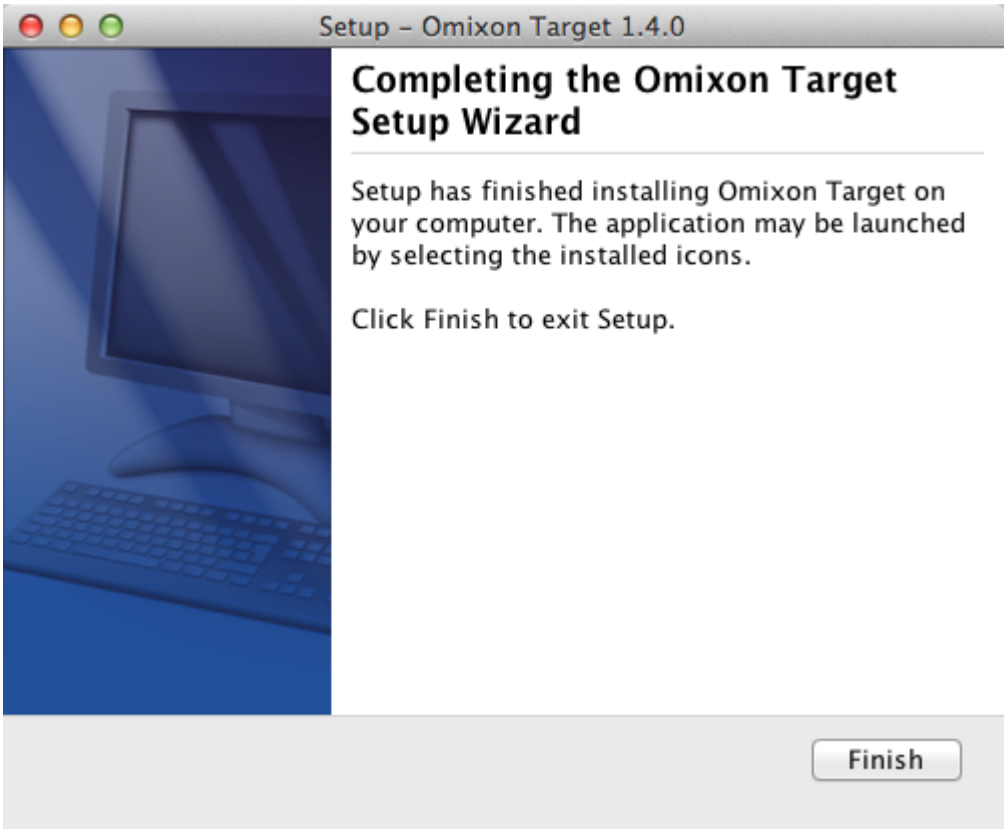


Select temporary files directory.

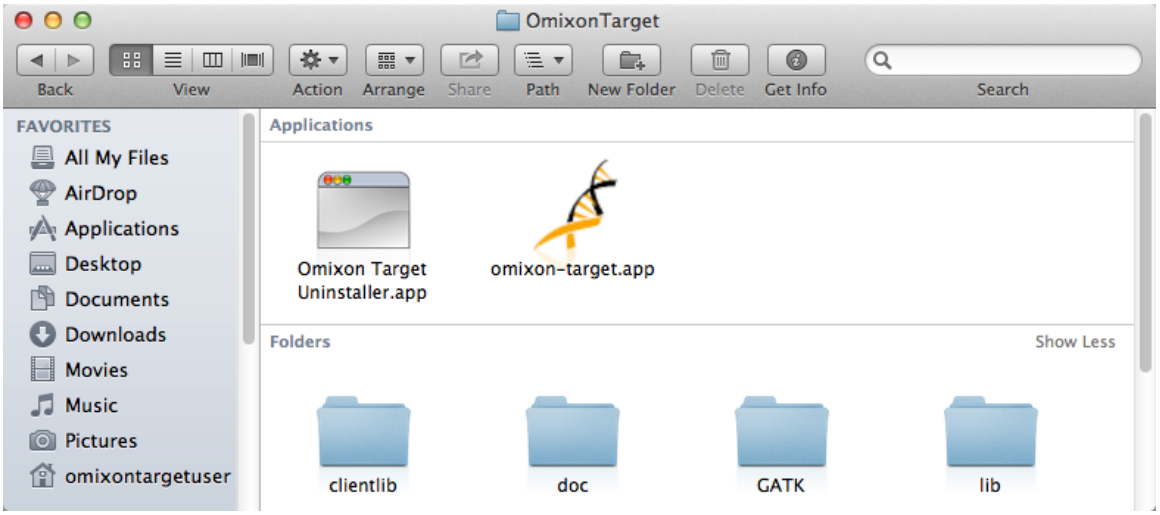


After clicking the Next button, the installation process starts.

When it is finished, the Completing the Omixon Target Setup dialog indicates the success of the installation.



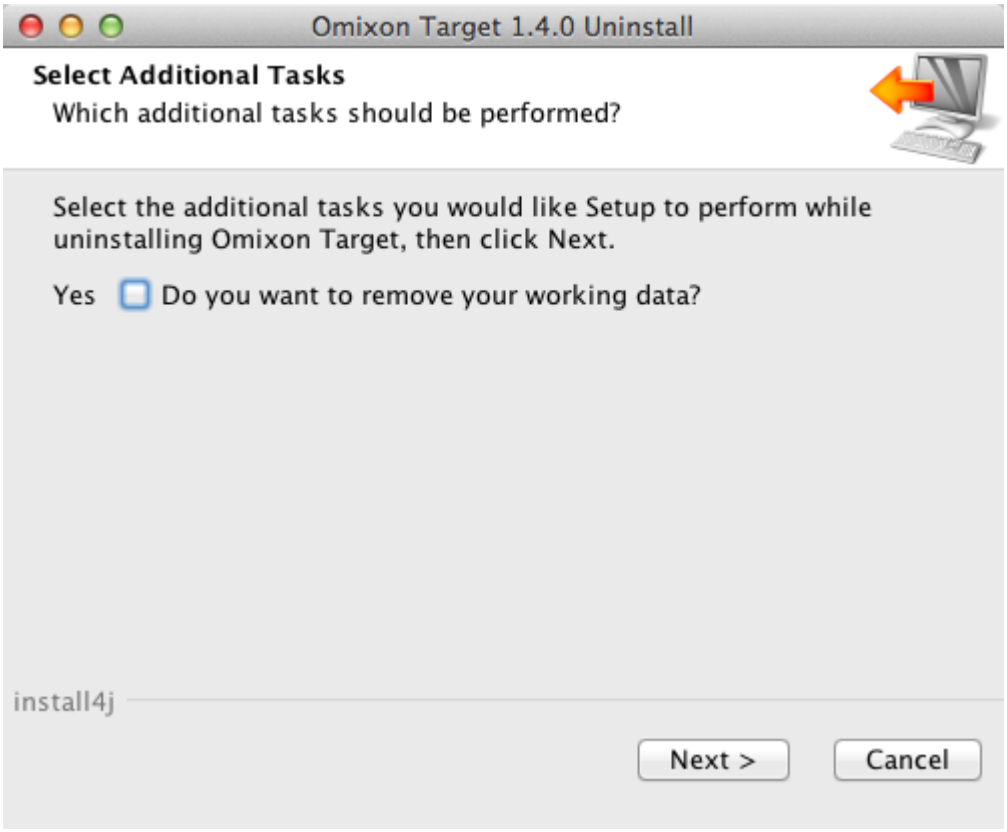
In Application list, 'Omixon Target' is available now. By clicking the 'Omixon-Target.app' icon, application starts and is ready for use.



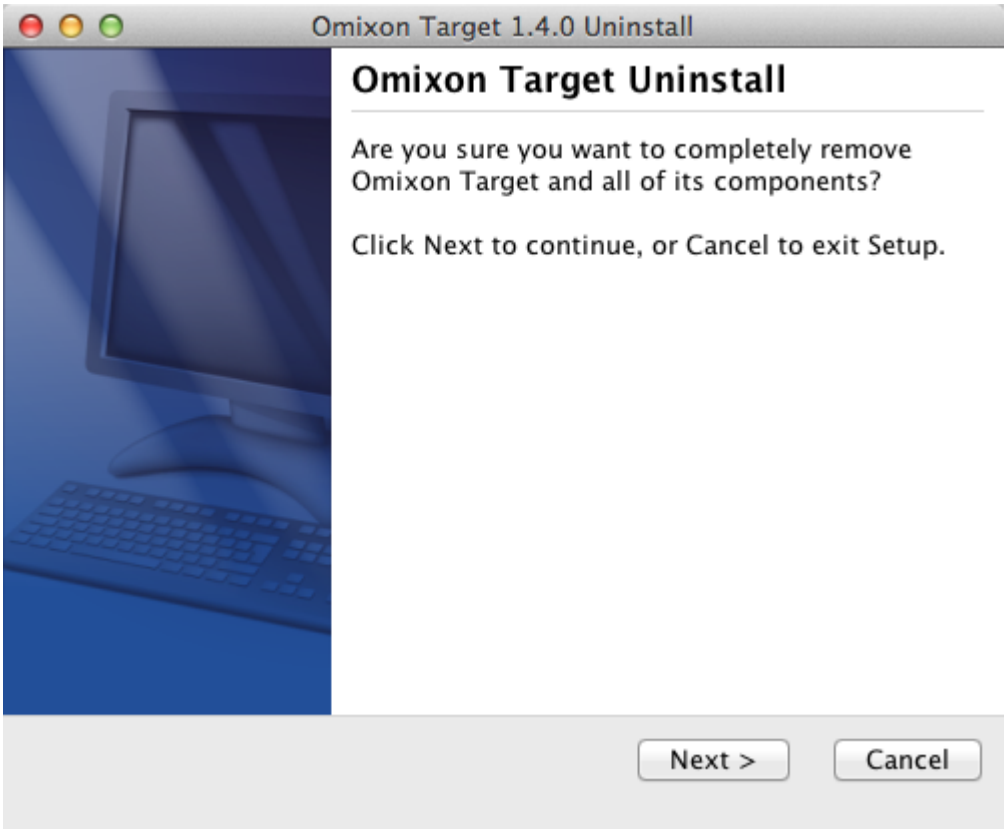
Uninstalling the product

If you want to remove Omixon Target, run the uninstaller app which is available in the directory where the application was installed.

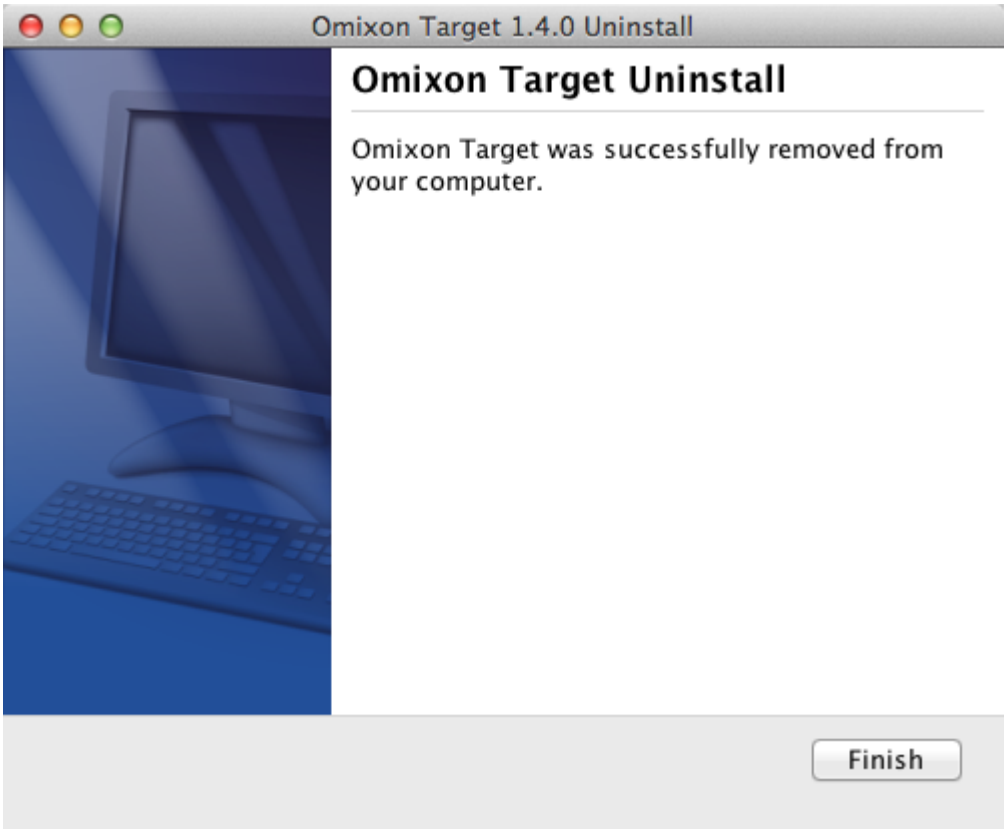
If you intend to use future version of Omixon Target and you don't want your work data to be deleted, keep the checkbox unchecked. If you wish to cleanup the working directory, then selet the option 'Yes'.



Click 'Next' to proceed with the uninstallation process.



Finally, the installer informs you about the outcome of uninstall.



Linux installation guide

We provide two versions of installer package for Linux operating systems, both are bundled with a Java Runtime Environment (JRE):

- 64 bit version, named omixon-target_unix_1_4_0_with_jre.sh;
- 32 bit version, named omixon-target_unix_1_4_0_x86_with_jre.sh.

The install packages are single file shell scripts, suitable to install on various Linux distributions.

The 32 bit operating systems are not officially supported (not fully tested), but that does not mean that the application cannot run. The current limitation of 32 bit operating systems is mostly related to the amount of memory available.

The downloaded installer does not have the permissions to run directly. Open a terminal window to make it executable with the following command:

```
chmod +x omixon-target_unix_1_4_0_with_jre.sh
```

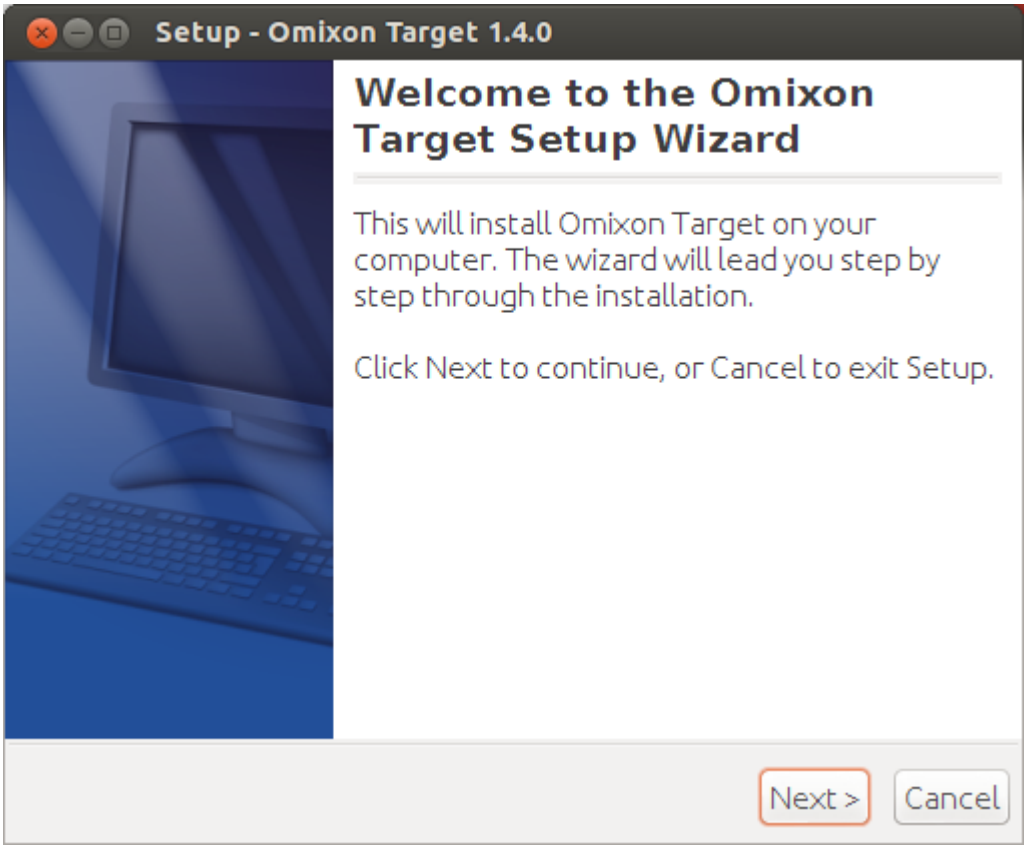
After that, it can be started with the following command:

```
./omixon-target_unix_1_4_0_with_jre.sh
```

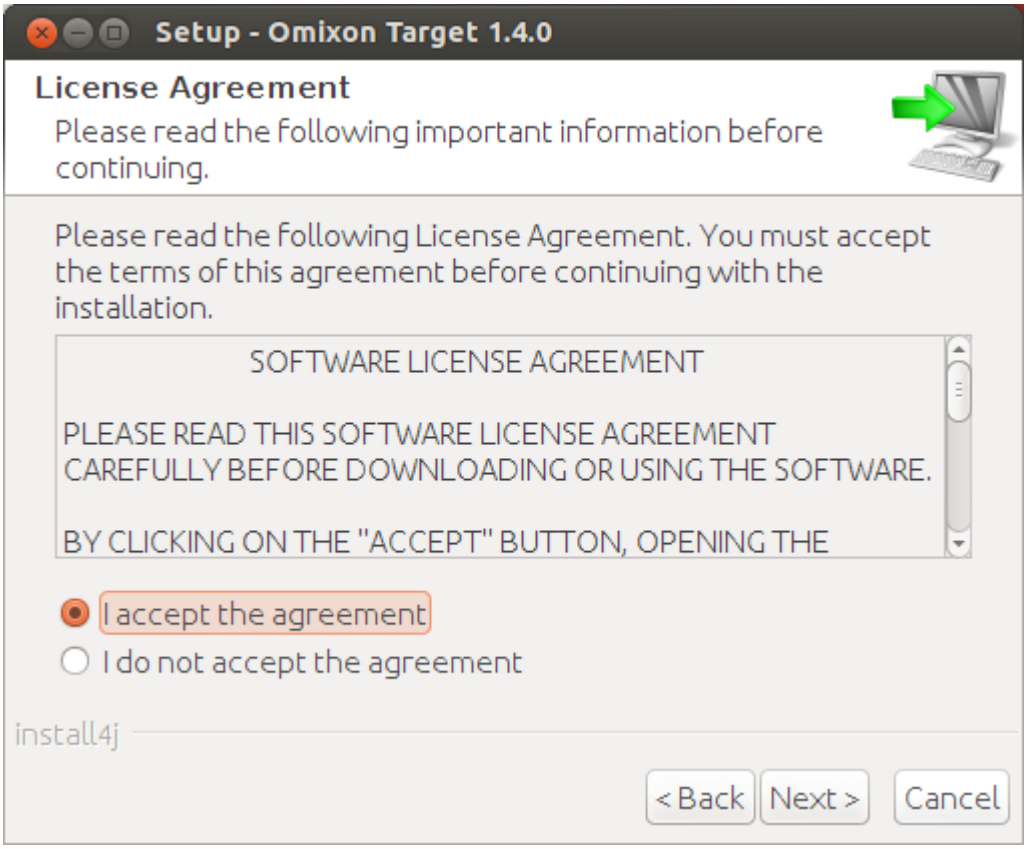
Installation steps

Launch the installer shell script.

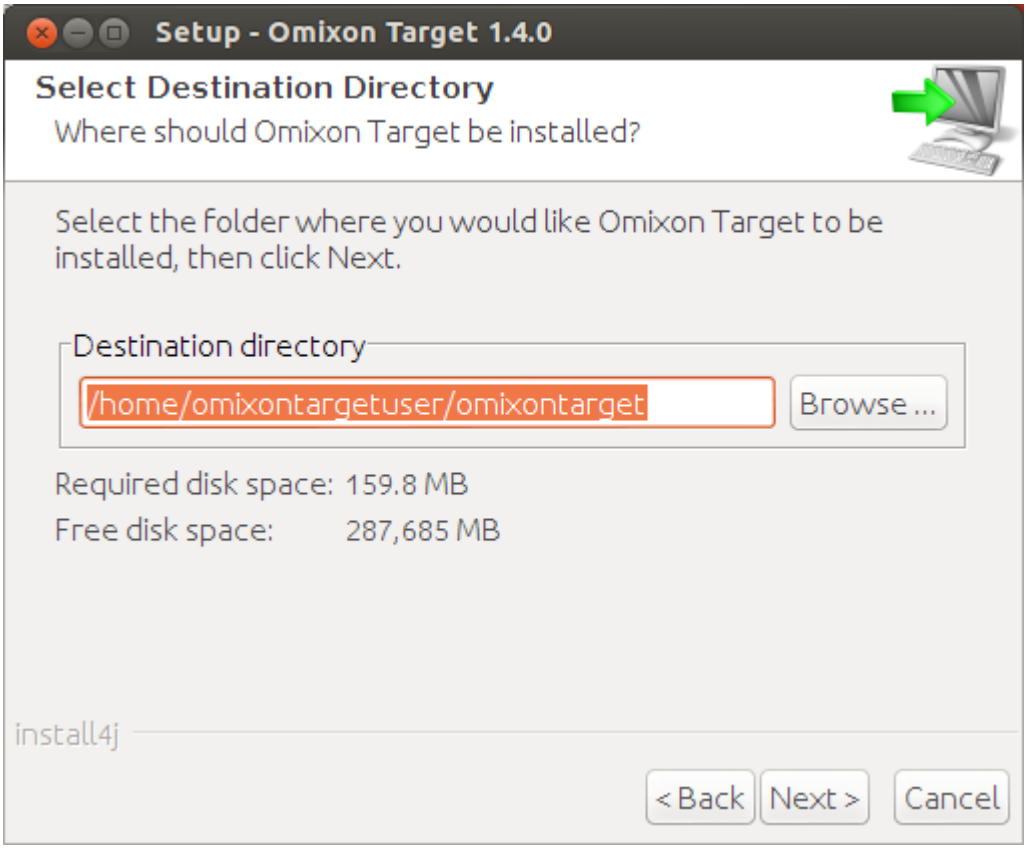
After launching the installer, Welcome dialog of the setup wizard appears.



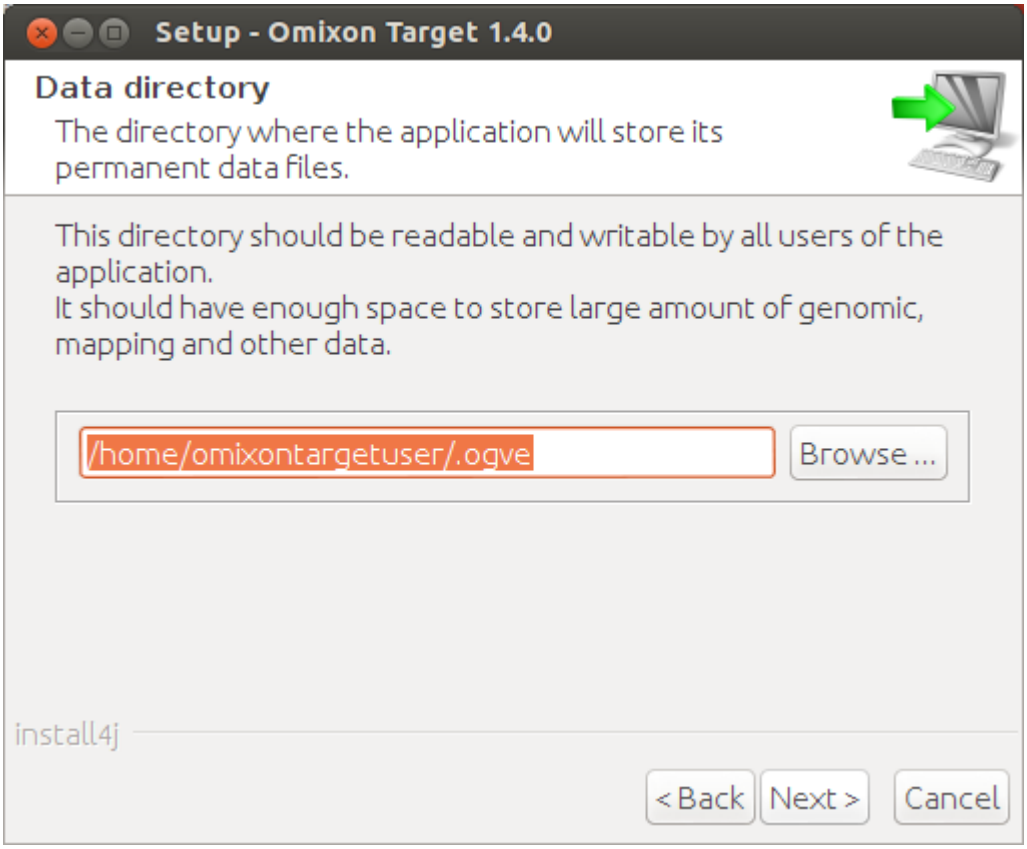
Please, accept the software license agreement to proceed with the installation.



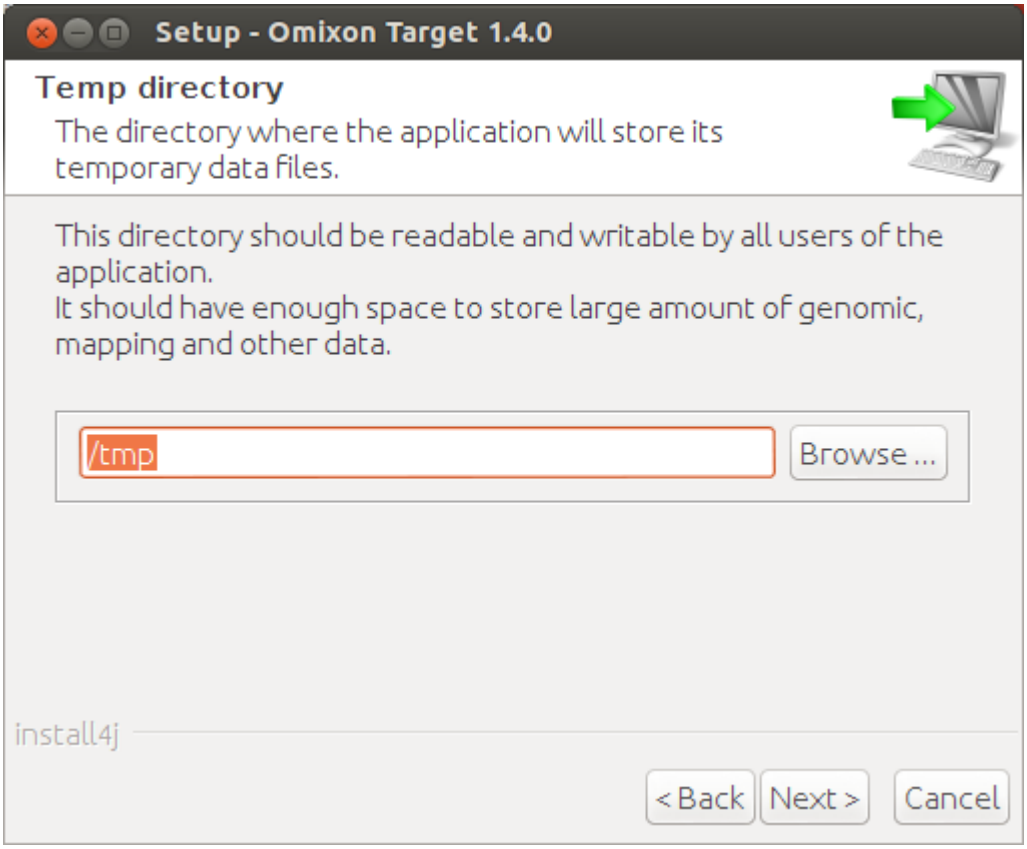
Select the path where you would like the application to be installed.



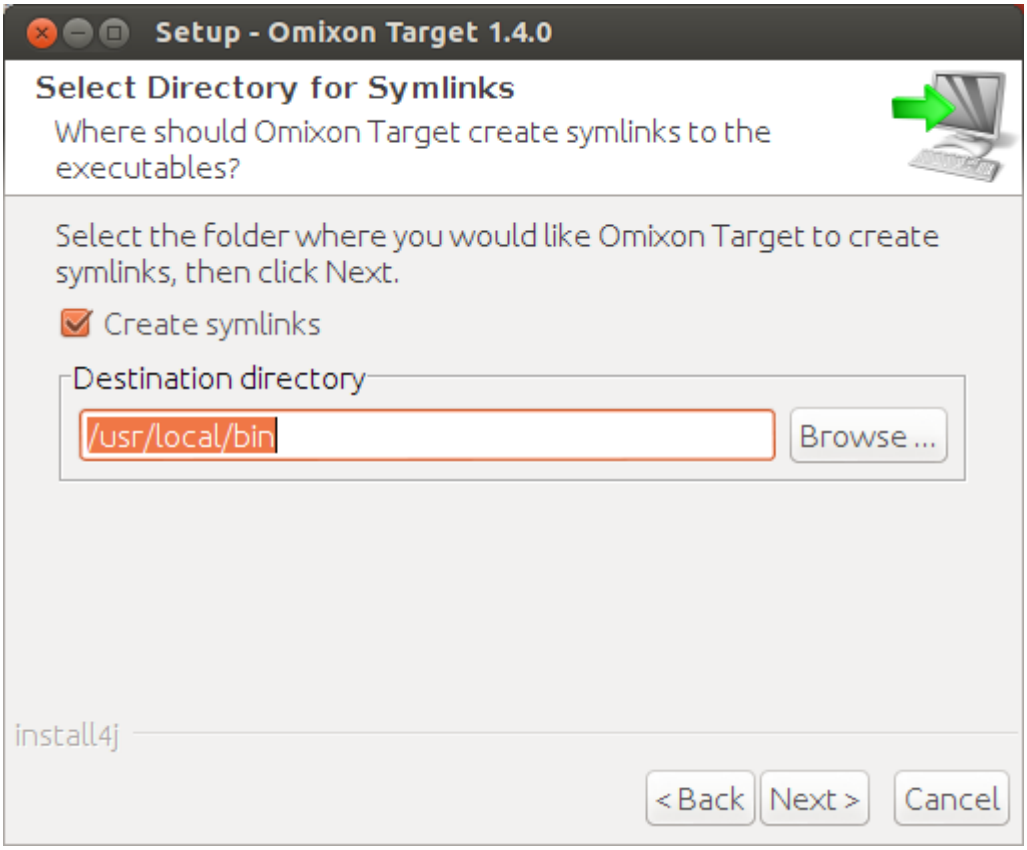
Select data directory where the permanent data files will be stored.



Select temporary files directory.

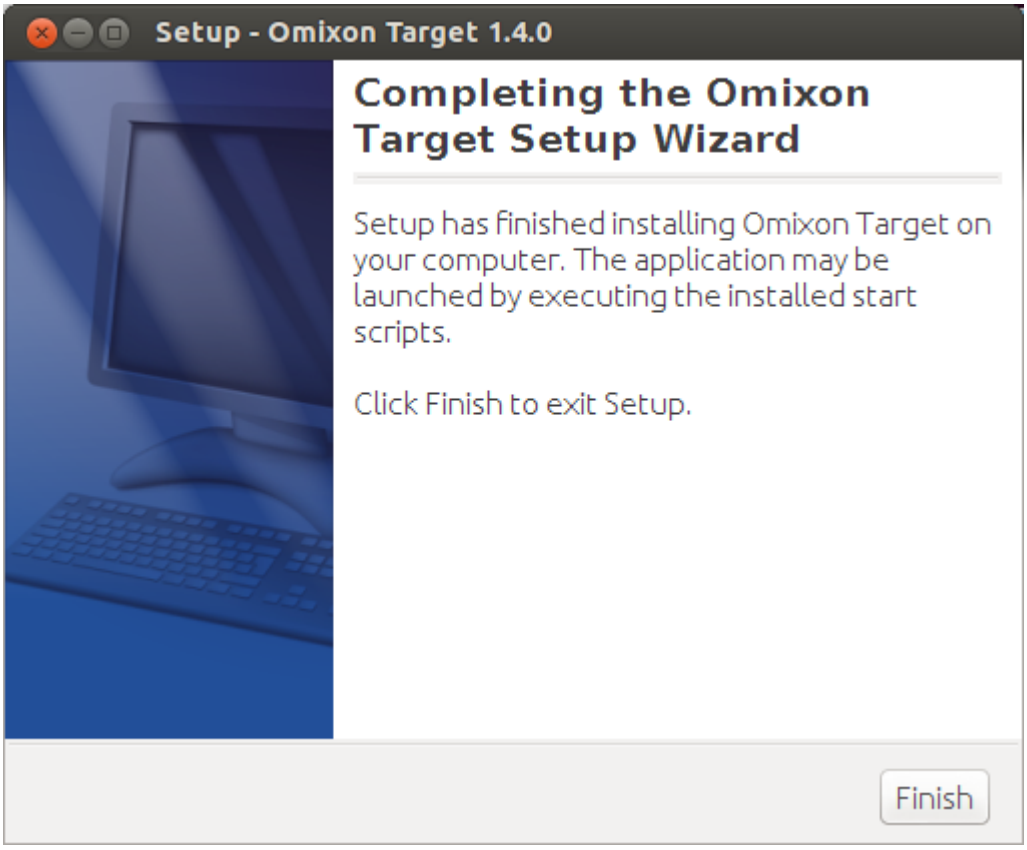


Set symlink location. It is useful for users who start the application from the command line. Since the installer is creating a startup icon in the System menu, this symlink is optional for most users.

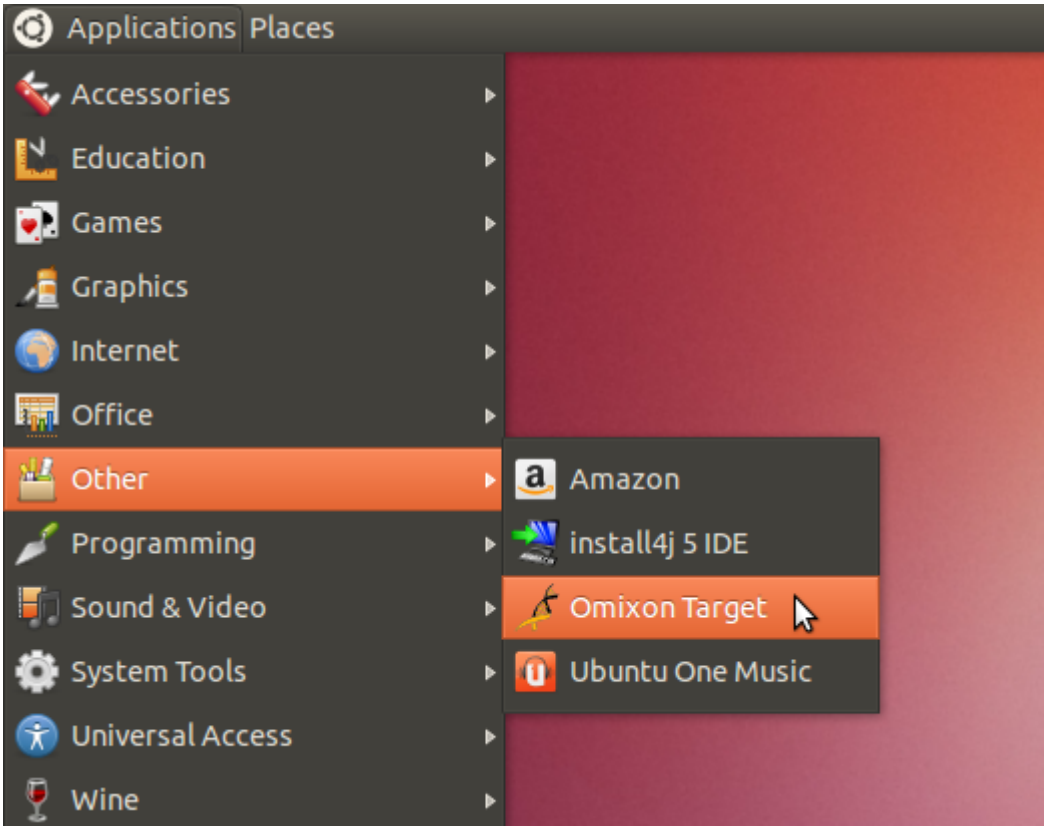


After clicking the Next button, the installation process starts.

When it is finished, the Completing the Omixon Target Setup dialog indicates the success of the installation.



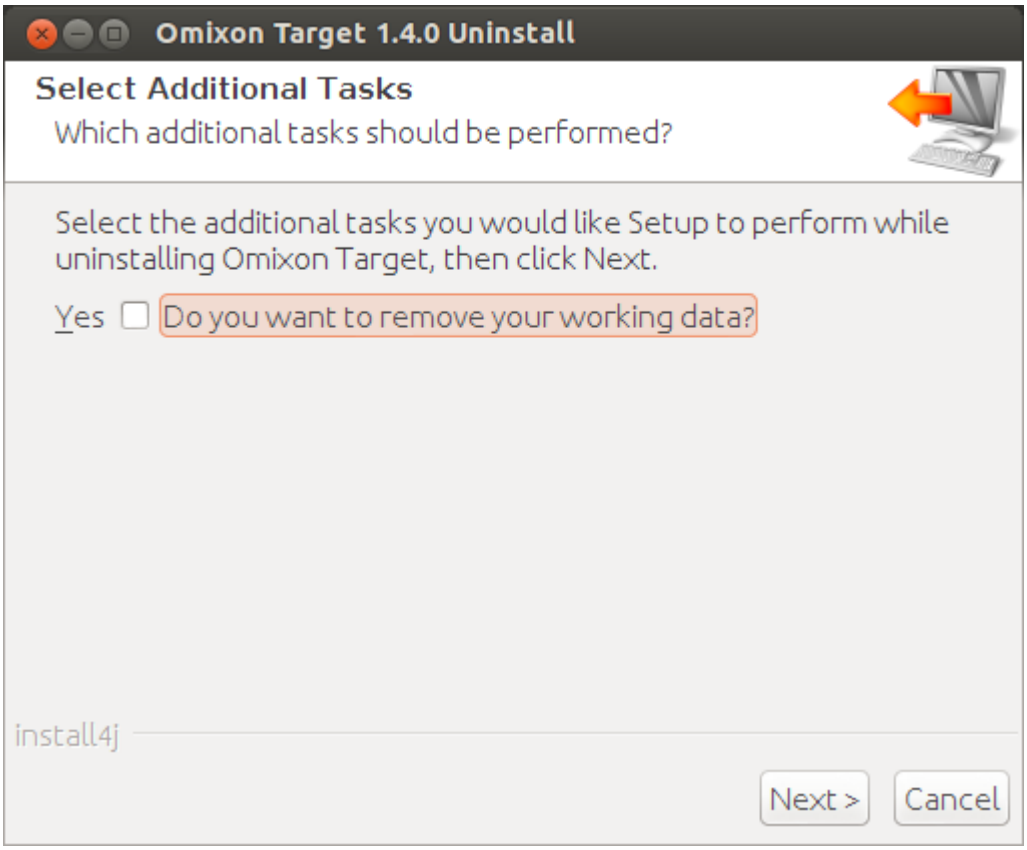
An application icon for Omixon Target is placed into the System menu. By clicking on this icon, the Omixon Target application starts and is ready for use.



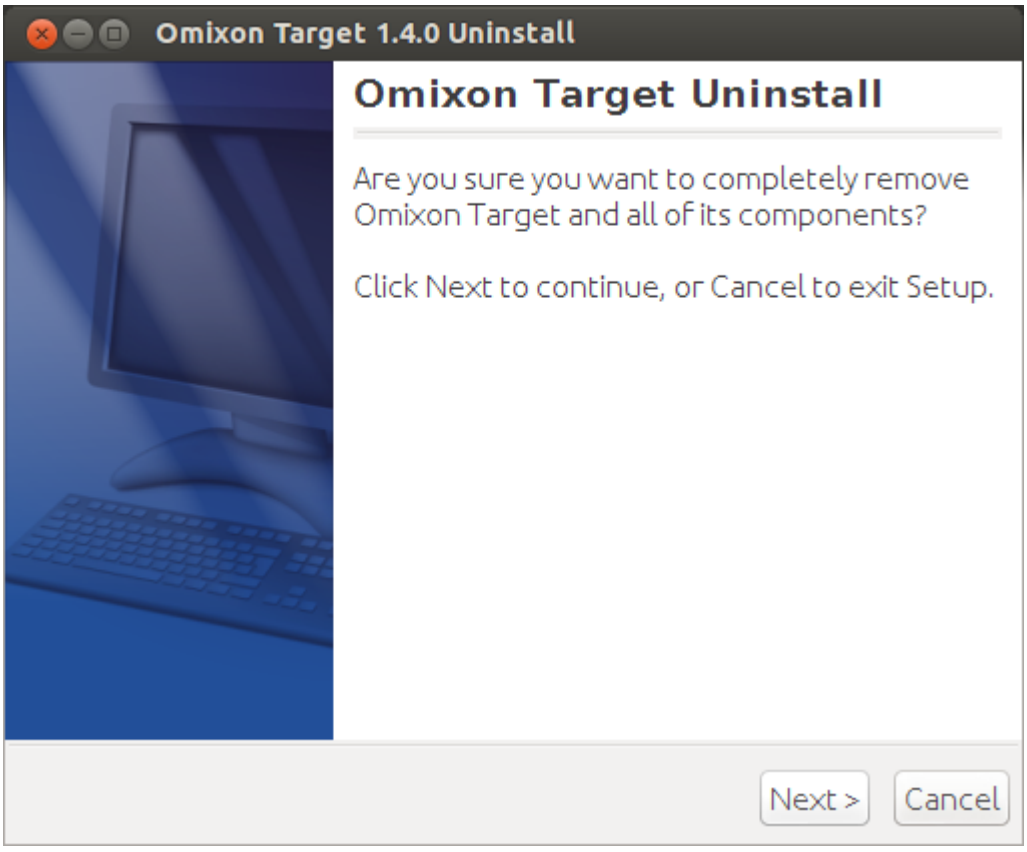
Uninstalling the product

If you want to remove Omixon Target, an uninstaller script is available in the directory where the application was installed.

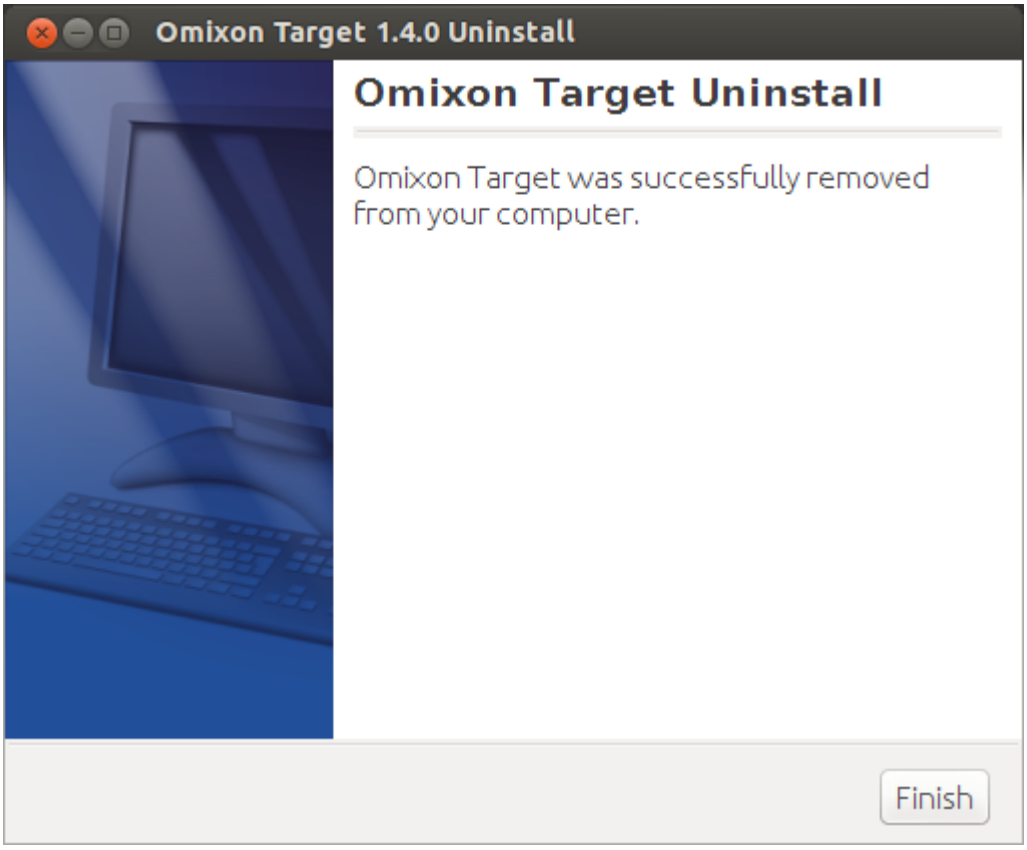
If you intend to use future version of Omixon Target and you don't want your work data to be deleted, keep the checkbox unchecked. If you wish to cleanup the working directory, then selet the option 'Yes'.



Click 'Next' to proceed with the uninstallation process.



Finally, the installer informs you about the outcome of uninstall.



This symlink is useful for users who usually start the application from the command line. Since the installer is creating a startup icon in the system menu, this symlink is optional for most users.

Pro Data Analysis Module installation notes

Introduction

The Pro Data Analysis Module is an optional module which supports additional features compared to the Standard Data Analysis Module.

The Pro Module is supported only on a 64 bit Linux platform.

The Pro Module uses several well-known bioinformatics tools behind the scenes. You need to install these tools manually. Once they are on the PATH, they can be used.

General notes and configuration

In case you have all the tools installed already and they are available on your PATH then no additional configuration is needed. Otherwise you have to configure them manually for Omixon Target to let the application find them.

The configuration file is called **tools.properties** and located in the **proserver-natives/etc** folder of your installation directory. You can find a configuration line for each tool and you can provide the full path to the executable of the tools.

List of tools

Below is the list of tools to be installed in your environment to get all Pro features working.

Samtools

Homepage of the tool: http://samtools.sourceforge.net

Version used for testing: *0.1.8*

Installation with package manager on Debian based systems: *apt-get install samtools*

Bowtie

Homepage of the tool: http://bowtie-bio.sourceforge.net

Version used for testing: *0.12.8*

Installation with package manager on Debian based systems: *apt-get install bowtie*

FASTX-Toolkit

Homepage of the tool: http://hannonlab.cshl.edu/fastx_toolkit/

Version used for testing: *0.0.13.2*

Installation with package manager on Debian based systems: *apt-get install fastx-toolkit*

Seqtk

Homepage of the tool: https://github.com/lh3/seqtk

Revision used for testing: *771d60b8f774482a77d60dd8559d17bd487c56e8*

Installation from source:

> *git clone https://github.com/lh3/seqtk.git*
>
> *cd seqtk*
>
> *make*
>
> *cp -f seqtk /usr/local/bin*

Burrows-Wheeler Aligner

Homepage of the tool: http://bio-bwa.sourceforge.net

Version used fot testing: *0.6.2*

Installation with package manager on Debian based systems: *apt-get install bwa*

Omixon Target Server setup

There is a beta version of the Omixon Target Server available for testing. Contact sales@omixon.com if you would like to try this.

Omixon Target Server accepts multiple client connections simultaneously. It allows collaborative work and sharing of analysis results between users. It also allows 32 bit clients (Windows, Linux, Mac) to be used to connect to a 64 bit server (also Windows, Linux, Mac) or Pro server (Linux only).

Configuring the server takes only a few easy steps at first startup.

Installation

Omixon provide two installers for the server version:

- Server installer - containing the server and a single client;
- Client installer - containing a single client only.

The server is bundled together with a client, so there is no need to install a client separately on the server computer.

To run the server installer, system administrator rights are required. The server installer can also be run from the command line, if no monitor is directly attached (please contact support@omixon.com for details). The client-only installer doesn't need system administrator rights.

The steps of the server and client Setup Wizard are equivalent to the desktop version's. For both, please, follow the desktop Installation Guide of your operating system found in the preceding sections of this manual.

Configuring the server

After server installation has finished, edit *omixon-target-server.vmoptions* in Omixon Target application folder to set host and port parameters.

> *IMPORTANT*
>
> *The last line in* .omixon-target-server.vmoptions *must be followed by a line feed.*

The default parameters are:

-Domixon.server.host=localhost
-Domixon.server.port=4380

Host

Modify 'localhost' value. Add hostname (e.g.: target-server) or full domain name (e.g.: target-server.mycompany.com) or IP address.

> *IMPORTANT*
>
> *In the client's connection configuration, 'Server host' must be set the exact same value as '-Domixon.server.host' in the server configuration.*
>
> *Make sure that the server hostname resolves to the same IP on the server and client machines otherwise the clients will get a connection refused error. It is a common configuration problem that the hostname resolves to a different IP on the server (e.g. to 127.0.0.1 via an internal network interface) which causes connection deny.*
>
> *It is always safe to specify an exact IP address both for the server and clients which is reachable from all the related machines across the network.*

Port

Leaving the default port setting is recommended.

Starting the server

Linux

Start the server daemon from the command line with one of the following input arguments. The available arguments get listed when the server is launched without any.

Usage: omixon-target-server {start|stop|status|restart|force-reload}

start	stop	status	restart or force-reload
Starts server if status is stopped.	Shuts server down if status is running.	Returns running status.	Restarts server or starts if isn't running.

Automatic startup is not set for Omixon Target Server. To set up automatic startup and specify in which runlevel to start the service, use a service configuration utility like *chkconfig* or *update-rc.d*.

Windows

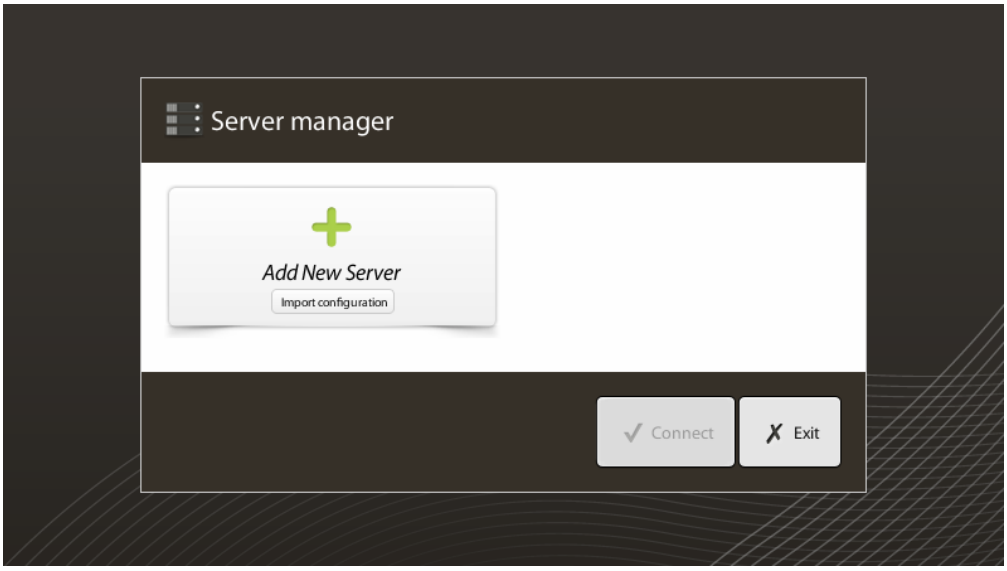
To start, stop and set startup type for Omixon Target Server, launch 'Control panel - Administrative tools - Services' and modify 'Omixon Target server' service properties. By default, startup type is set to automatic and the service is run under the System user account.

MAC OS X

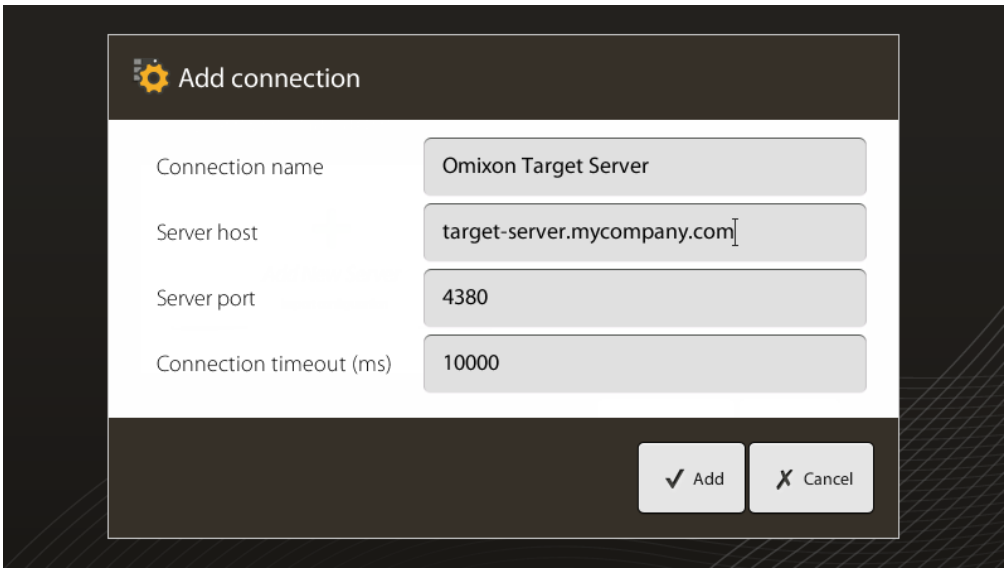
Start and stop the server the same way as on Linux.

Connecting the client

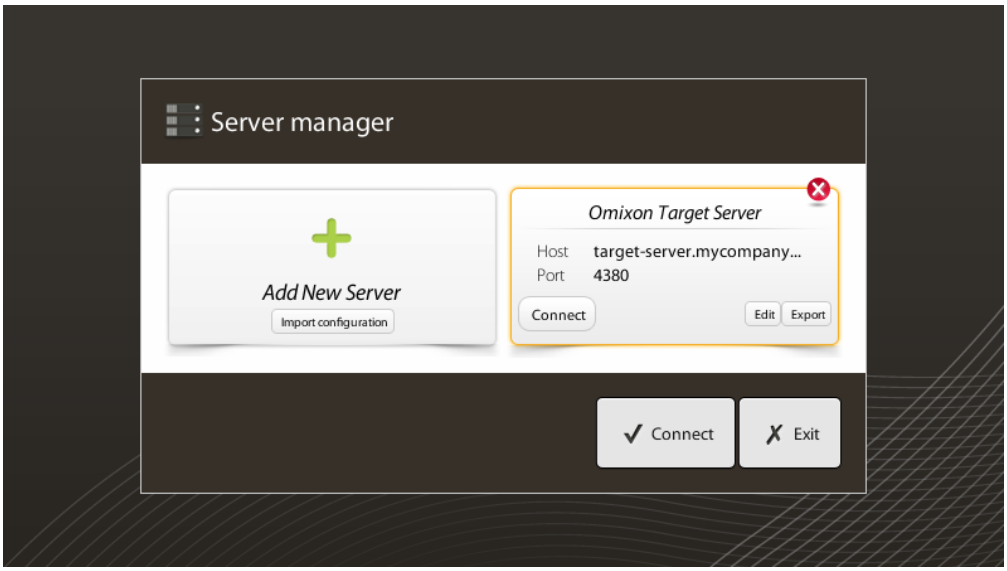
Launch client application. On 'Server Manager' screen, select 'Add New Server' to setup client-server connection.



Name your connection and enter exactly the same host and port settings which were set for server.



Select server connection and do connect.



> *IMPORTANT*
>
> *Check your firewall settings on the server computer. Setup your firewall to let Omixon Target Server accept incoming connections.*

The super user role

The very first user on the server becomes a 'super user', a server administrator who manages other user accounts. For this reason the system administrator is supposed to register and login to Omixon Target at the setup period. Thenceforward registration is not available any more and the login screen welcomes all connecting clients. The clients' user accounts are generated and given by the super user.

Exporting and importing connection configuration

Instead of configuring the connection manually, the users can choose to import a configuration file by clicking 'Import configuration' button on 'Add New Server' card in 'Server Manager'. The file is provided by the system administrator who needs to export the connection settings into a file by clicking 'Export' on the selected connection in 'Server Manager'.

Acknowledgements

Collaborators

We would particularly like to thank SmartArt, who prepared all the graphics and images used within the tool.

Third Party tools

The Genome Analysis Toolkit (GATK) from the Broad Institute is used for the variant calling - an older version (1.6x) is used.

A number of tools from SamTools Picard are also used within the tool, for handling and manipulating SAM and BAM files.

The Pro Data Analysis Module allows access to a number of commonly available, open source bioinformatics tools. These are not however bundled within the Omixon Target application.

The IMGT/HLA database is used for the HLA typing, here are the citations:

- Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SGE
  The IMGT/HLA Database
  Nucleic Acids Research (2011) **39 Suppl 1**:D1171-6
- Robinson J, Malik A, Parham P, Bodmer JG, Marsh SGE:
  IMGT/HLA - a sequence database for the human major histocompatibility complex
  Tissue Antigens (2000), **55**:280-287

Burrows-wheeler aligner software package for mapping low-divergent sequences against a large reference genome:

- The short read alignment component (bwa-short) has been published:
  Li H. and Durbin R. (2009)
  Fast and accurate short read alignment with Burrows-Wheeler Transform.
  Bioinformatics, 25:1754-60. [PMID: 19451168]

- BWA-SW:
  Li H. and Durbin R. (2010)
  Fast and accurate long-read alignment with Burrows-Wheeler Transform.
  Bioinformatics, Epub. [PMID: 20080505]

SnpEff variant annotation and effect prediction tool:

> Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM.
> A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.
> Fly (Austin). 2012 Apr-Jun;6(2):80-92. PMID: 22728672 [PubMed - in process]

Tutorials

First Use

The easiest way to get started with Omixon Target is to choose a pre-configured Profile to work with (or more than one). Each Profile already includes a reference and a target.

Fast Start:

- Add Profile
- Import sequencing (fastq) data
- Map the reads and call variants
- Analyse the results

This tutorial only explains the steps required for the Fast Start. There are other tutorials available for using the Expert profile and setting the tool up manually.

Adding a Profile

This is the first use, fast start tutorial for the **Data Analysis** module.

The first step of the Fast Start involves adding a Profile (the alternative is to create your own 'custom' Profile using Create Profile - see the next tutorial for details).

This function is available from the Data Analysis Dashboard. Adding a new Profile will start a new background task, which (by default) will download and install all the data required to use the Profile. It will also create a new Experiment and Analysis for you, which will be pre-configured with the sequencer chosen, and with the reference and target from within the Profile.

After this Add Profile task has finished (it will take 3 to 10 minutes, depending on the speed of your internet connection), you can move to step 2 of this Fast Start tutorial.

Download problems

If for some reason you cannot download files from within the tool (e.g. due to network security settings) you can download the reference files manually and use the 'Select Local file(s)' option instead of the 'Download file(s)' option in the wizard.

The files for the in-built profiles can be downloaded from here:

BRCA profile:

- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr13.zip
- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr17.zip

CFTR profile:

- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr7.zip

Data Analysis Example Data

Data Analysis is an optional Module built into Omixon Target.

We offer a few example datasets that go with the in-built profiles. These allow you to easily try out the 'Map and Align' and 'Call Variants' functions within the tool.

BRCA profile example data – paired illumina data (two fastq files)

http://omixon-download.s3.amazonaws.com/target_brca_example.zip

In order to import the fastq you should use the 'import multiple samples' function from the Analysis Dashboard (select 'Go' on the appropriate Analysis, after Add Profile). For the paired data you need to select the first file (.1.fastq) in the first file selector in the wizard, and the second file (.2.fastq) in the 'paired data' file selector. The default pair parameters (orientation, distance) are fine for the example data sets.

Once you have imported the fastq, you will be given the option to run the Map and Align function for the Sample.

Import sequencing (fastq) data

Once the Add Profile task from step 1 has finished, you can start to work with the Profile.

The first step is to select a Sample to work with. There is a small data management hierarchy within the tool - the Samples are grouped together with an Analysis, and the Analyses are grouped together within an Experiment.

Select the example Experiment by clicking on it (which will navigate to the Experiment Dashboard), and select the example Analysis in the same way (moves to the Analysis Dashboard). You can click on the Create Sample button in the Analysis Dashboard and create a brand new Sample to work with.

- Select the Sample you wish to work with by clicking on it (moves to the Sample Dashboard). Select the 'Import sequencing data' button, and choose a fastq or sff file (or a pair of files, for paired data) from the file system. Click Finish. This will start a background task. Once this task is finished move to the next step of the tutorial.

Map the reads and call variants

Once you have some fastq data imported, then you can map and align this data plus call variants in a single function.

Either:

- Select a Sample with fastq data and from the Sample Dashboard select 'Map and Align' from the buttons on the left or
- Select a Sample (or more than one) from the list of Samples in the Analysis Dashboard and select 'Map and Align' from the buttons on the top.

This will start the Map and Align wizard. You shouldn't need to change any settings in here - you can review the settings and when ready press Finish.

This will start one background task per Sample. Once the background tasks are finished go to the next step.

Results - analysis and visualisation

Once the 'Map and Align' (including variant call) is finished, you will now be able to start analysing the results. You should see at least two new data items within the Sample - one for the mapped data (in BAM format) and one (for each reference contig) for the variant calls (in VCF format).

You can visualise the results in the Genome Browser. From the Sample Dashboard select 'Genome browser', or from the Analysis Dashboard select the Sample and then choose 'Browse Samples'. This will open the Genome Browser centered on the target you have defined in the Analysis/Experiment.

You can start to analyse the variants. From the Sample Dashboard select 'Analyse sample', which will start a new dashboard-style screen with the same name. In here you can the variants that have been called, in lists that are split by on-target and off-target lists. You can select items in these lists and then 'Browse Feature' which will switch to the Genome Browser and show you the variant and short read reads at that position.

From here you can also start the Approval Dashboard (in order to approve or reject variants for further analysis), or you can start the Report Dashboard (to view summary reports for this Sample).

Custom Profile

If none of the built-in Profiles are useful, then you will need to create your own 'custom' Profile within the tool.

The first thing to note is that the Profiles are essentially just convenience wrappers around some Reference sequences and a Target definition. So to create the equivalent of a new Profile all you need to do is to import some reference data (and possibly some known variants for that data), and then create a new Target based on that reference data. Even having a Target is not mandatory, but using one is recommended as there are lots of 'Target-based' functions within the tool.

This tutorial will walk you through the steps needed to do this, with an example.

In this example we will create the profile for one of the 'ready to use' Ion AmpliSeq panels.

Ion AmpliSeq™ Comprehensive Cancer Panel list of target genes (.csv):
http://tools.invitrogen.com/downloads/cms_103573.csv (from page
http://products.invitrogen.com/ivgn/product/4477685?ICID=search-product)

Here is a description of a run of the kit, with some test data, if you would like to try it:
http://ioncommunity.lifetechnologies.com/docs/DOC-3043

The amplicons in the kit are documented in a bed file:
http://omixon-download.s3.amazonaws.com/AmpliSeqCancerAmplicons.bed

Option 1: Using the 'Create HG19 Profile' function (recommended)

The fastest way to create a new Profile is to use the one of the two Create Profile functions from the main Data Analysis dashboard. These wizards wrap up the a number of the other steps explained below. There are two options for Create Profile - creating one based on the HG19 reference genome, or creating a 'custom' one with another reference genome.

For the AmpliSeq example, you can do this very simply by using the Create HG19 profile function. This will automatically download and extract all the required reference chromosome data from the Omixon web site.

You should start at the Data Analysis Dashboard.

- Click on 'Create HG19 Profile'
- Fill in the details - Profile Name (e.g. AmpliSeq), Target Name (e.g. AmpliSeq Amplicons), Experiment Name (e.g. 'All my AmpliSeq data'), Analysis Name (e.g. 'First go'). Press Next.
- Choose one or more sequencers for your Analysis (tip - use CTRL or SHIFT click to select multiple items). Press Next.
- Select Annotation Type - Amplicon. Press Next.
- Select the AmpliSeqCancerAmplicons.bed file from the file system.
- Leave the next step set to 'Download file(s)'.
- Press Finish.

This will start a new background task. This task will take quite a lot of time - it will download most of the HG19 genome and install this reference data into the tool.

The task will parse the bed file and identify which chromosomes it needs from within the bed file. It is important that the bed file uses the standard HG19 notation for chromosome names, i.e. the name has to be 'chr1' and not just '1'. If the chromosome is already configured within the tool then it will not be downloaded again.

Once this task has completed you should see your new Experiment (and Analysis) and you can start importing sequencing data and using your new Target.

Download problems

If for some reason you cannot download files from within the tool (e.g. due to network security settings) you can download the files manually and use the 'Select Local file(s)' option instead of the 'Download file(s)' option in the wizard.

The pre-packaged HG19 zip files can be downloaded from here:

- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr1.zip
- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr2.zip
- etc... replace the chr* number with the chromosome you would like to download

Option 2: Using the 'Create Custom Profile' function

If you would like to create a new Profile not based on the HG19 reference genome then this wizard is the easiest way to achieve that.

This wizard is very similar to the Create HG19 Profile wizard above. The main difference is that you can either select an existing reference genome or create a new one during the wizard steps.

You can manually create a new reference genome (see steps below) before starting this wizard, if you wish.

This will start a new background task. Once this task has completed you should see your new Experiment (and Analysis) and you can start importing sequencing data and using your new Target.

Option 3: Manually Creating a Profile

You can also manually recreate the Create Custom Profile steps using the individual functions.

Create a Reference

You should start at the Data Analysis Dashboard.

- Click on 'References' and then click on 'Create New Reference'
- Select the name for your reference and species
- Press 'Finish'

Importing Reference Data

Omixon Target needs individual chromosomes to be added (the whole Human Genome multifasta file will not work correctly within the tool).

You should start at the Data Analysis Dashboard.

- Click on 'References' and then click on the new custom reference you just created in the list.
- Select the 'Import reference data' item in the 'Actions' menu on the left.
- Choose the reference chromosomes - you can select multiple files from the list in one go - e.g. chr1.fa, chr2.fa

This will start a new background task. Once this task has completed you can proceed to the next step.

Importing Known Variants

This is an optional step, but having known variants within the tool can assist with the Variant Calling algorithms.

If you are still in the Reference Genome Dashboard then you are starting in the right place. Otherwise start at the Data Analysis Dashboard, click on 'References' and then click on the new custom reference in the list.

- Select the 'Import known variants' item in the 'Actions' menu on the left.
- Select the matching vcf files for the chromosomes. Note that we need the name of the chromosome (e.g. chr1) to appear in the name of the matching .vcf file - here is an example of one we use within one of the Profiles: dbsnp-hg19-chr1.vcf.

This will start a new background task. Once this task has completed you can proceed to the next step.

Import Target

You should start at the Data Analysis Dashboard.

- Click on 'Targets' and choose the 'Import Target' button from the menu above the list of Targets.
- Give your new Target a name, e.g. Custom Cancer Gene Panel
- Select the Annotation Type - Amplicon
- Select the bed file for your target regions from the file system.
- Choose the Reference - your new custom reference
- Select the Contigs (chromosomes)
- Press Finish

This will start a new background task. Once this task has completed you can create a new Experiment and start using your new Target.

New Target

The first thing to mention is that you don't **have** to use a Target. However, a number of the features within the tool are built to support the Target concept, and having a well-defined target can lead to better analysis results.

There are two ways to create a new Target within the tool. The first (and easiest) is to import a set of Target annotations using a .gff or .bed file. The second way is to manually import annotations for a reference, and then manually create a Target to use those annotations.

Import Target

You should start at the Data Analysis Dashboard.

- Click on 'Targets' and choose the 'Import Target' button from the menu above the list of Targets.
- Give your new Target a name
- Select the Annotation Type - Amplicon, Exon or Gene
- Select the .bed or .gff file from the file system.
- Choose the Reference
- Select the Contigs
- Press Finish

This will start a new background task.

Manual Target Creation

There are three steps involved with creating a new Target, the first of these is actually performed with the Reference Genome that will be

used. For the purposes of this tutorial, it is assumed that the Reference Genome has already been created and configured correctly (see the 'New Reference' tutorial for more on this).

Reference Genome configuration (via the Reference Genome Dashboard):

- Import Annotations (Genes or Exons)

Target configuration (via Targets lists and Target Dashboard)

- Create Target
- Configure Target

Adjusting Memory Usage

When you install Omixon Target, it comes preconfigured with some memory settings. These will vary, depending on which version of Omixon Target you have chosen to install. For the 64 bit versions, the memory is set to 5GB, for the 32 bit versions it is set to 1200MB.

For most analysis, the maximum 5GB memory is enough. However, it might not be sufficient for the underlying alignment algorithms to run. If you get 'Out of Memory' or 'GC Limit Exceeded' errors, then you will need to:

- either increase the amount of memory available to Omixon Target
- or run the tool with reduced sensitivity.

The first option is preferable. Running with reduced sensitivity is possible, and saves lots of memory, but doesn't give such good results.

Increase the amount of memory available to Omixon Target

This can be done by editing one of the Omixon Target configuration files. In your Omixon Target installation directory on Windows and Linux, there is a file called omixon-target.vmoptions, on the Mac it is found in /Applications/OmixonTarget.app/Contents/Info.plist . Open this file to modify a single VM ('virtual machine') option:

-Xmx5g

Replace original to the following:

-Xmx6g

Saving the file and restarting Omixon Target will make 6GB of memory available instead of the standard 5GB.

Run the tool with reduced sensitivity

If you are running a 32 bit version, and/or you don't have more than 5GB of memory available, the other option is to reduce the sensitivity of the mapping tool. You need to create a text file and to add the following line:

orm.sampling.default=4

Save it, then load it as a 'Parameters file' by using the 'Advanced Parameters' tab of the 'Map and Align' wizards. This will cause the mapping tool to sample every 4th position in the reference, rather than every single position. This saves a lot of memory - 60 to 70% of the total. You can also try a smaller value, such as:

orm.sampling.default=2

This will save about 30-40% of the memory required.

Overview

Home Dashboard

Introduction

This is main 'home page' for Omixon Target.

There are only four functions available from this screen:

- Data analysis
- HLA Typing
- Sandbox Genome Browser
- Settings

The **Data Analysis** module is for general analysis of sequencing targets, where sequencing data can be mapped and aligned against reference sequences, and variants can be called, visualised and analysed. It's an optional module, with a separate license.

The **HLA Typing** module is only for determining and visualising the HLA types within a set of NGS sequencing data. It is simple, fast and accurate. It's an optional module, with a separate license.

The **Sandbox Genome Browser** module is for quick visualisation of alignments, variants and annotations and for simple data analysis tasks (e.g. visual comparison of multiple data sets). The Sandbox is a free module.

The **Settings** function allows you to manage users, licenses and general settings within the tool, plus reset the tool to an 'empty' starting state.

Global Shortcuts

The following generic keyboard shortcuts are available in Omixon Target.

- **F1** key opens the Help for the current page or wizard.
- **F8** key closes the window and exits the application.
- **F11** key switches to fullscreen display mode to utilize as much space as possible for visualization - note that certain platforms and window managers are not working properly together with this function, the first time you use you will get a warning message about this.
- **ALT+F4** closes Omixon Target.
- **CTRL + C/V/X** copies/pastes/deletes text in text boxes.
- **SHIFT/CTRL + left click** can be used for multiple selection/deselection.

Data Analysis

Data Analysis Dashboard

The Data Analysis Dashboard is the starting point for general genomic analysis tasks

The recommended way to start doing data analysis is to use one of the in-built Profiles.

The tool starts with an 'empty' Expert Profile. The first step is either to manually configure this Profile, to add one of the existing pre-configured Profiles to the tool, or to create your own Profile(s). There are two pre-configured profiles available:

- BRCA
- CFTR

The pre-configured Profile to be used can be chosen with the 'Add Profile' function. This will start a new background task, which will download and import all the data required to configure the tool for the chosen profile.

In order to create a custom profile, the 'Create HG19 Profile' or the 'Create Custom Profile' functions can be used. With the 'Create HG19 Profile', only a single annotation file (in bed or gff format) is needed in order to configure the profile, the reference data and the known variants will be downloaded and imported automatically. The 'Create Custom Profile' makes it easy to create a profile with a custom (e.g. HG18 or non-human) reference, as it merges three of the other functions within the tool into a single wizard, for convenience ('Import Reference Data', 'Import Known Variants' and 'Import Target').

Once a Profile has been added or created, you can get started by selecting one of the Experiments listed in this Dashboard.

The targets and references for the whole tool can also be configured from here.

References and Targets

Add Profile

The tool starts up in '**Expert** mode', with only the Expert Profile configured and no reference data imported or targets set up.

A 'Profile' is simply a wrapper around a Target, plus a set of reference data and known variants.

It is possible to add a pre-configured Profile to the tool using the 'Add Profile' function.

The available pre-configured profiles are:

- **BRCA** (chr13 and chr17)
- **CFTR** (chr7)

The two pre-configured profiles can be set up automatically by using this function, which will also download and import all the data required for each Profile (reference sequences, known variants and snpEff database).

You can choose not to add a pre-configured Profile and create your own using the 'Create Profile' option. It's also possible to skip Profile set-up entirely, and simply configure the tool manually.

Download problems

If for some reason you cannot download the profile configuration file from within the tool (e.g. due to network security settings) you can download the necessary files manually and use the 'Select Local file(s)' option instead of the 'Download file(s)' option in the wizard.

For each profile, the setup files are packaged by chromosome and can be downloaded from the following links:

BRCA profile:

- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr13.zip
  http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr17.zip

CFTR profile:

- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr7.zip

The HG19 snpEff database can be downloaded from the following link:

- http://omixon-download.s3.amazonaws.com/target_snpeff_db_hg19.zip

Create HG19 Profile

The tool starts up in 'Expert mode', with only the Expert Profile configured and no reference data imported or targets set up.

A 'Profile' is a wrapper around a Target, plus a set of reference data and known variants.

A Target is a list of reference annotations, for either genes, exons or amplicons.

It is possible to add a pre-configured Profile to the tool using the 'Add Profile' function, or you can use this 'Create HG19 Profile' option instead. It's also possible to skip Profile set-up entirely, and configure the tool manually.

- Fill in the details - Profile Name, Target Name, Experiment Name, Analysis Name. A new Profile, Target, Experiment and Analysis will be created. Press Next.
- Choose one or more sequencers for your Analysis (tip - use CTRL or SHIFT click to select multiple items). Press Next.
- Select Annotation Type - Exon/Gene/Amplicon. Press Next.
- Select a .bed or .gff file from the file system, containing the annotations that describe your Target. Press Finish.

After completing the wizard, the chromosomes that are in the specified annotation files are downloaded automatically. For each chromosome, known variations and the snpEff database are also added.

Download problems

If for some reason you cannot download the profile configuration file from within the tool (e.g. due to network security settings) you can download the necessary files manually and use the 'Select Local file(s)' option instead of the 'Download file(s)' option in the wizard.

The setup files are packaged by chromosome and can be downloaded from the following links (for example):

Chromosome 13:

- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr13.zip

Chromosome 17:

- http://omixon-download.s3.amazonaws.com/target_ref_hg19_chr17.zip

etc.

Create Profile

The tool starts up in 'Expert mode', with only the Expert Profile configured and no reference data imported or targets set up.

A 'Profile' is simply a wrapper around a Target, plus a set of reference data and known variants.

A Target is simply a list of reference annotations, for either genes, exons or amplicons.

It is possible to add a pre-configured Profile to the tool using the 'Add Profile' function, or you can use this 'Create Profile' option instead. It's also possible to skip Profile set-up entirely, and simply configure the tool manually.

- Fill in the details - Profile Name, Target Name, Experiment Name, Analysis Name. A new Profile, Target, Experiment and Analysis will be created. Press Next.
- Choose one or more sequencers for your Analysis (tip - use CTRL or SHIFT click to select multiple items). Press Next.
- Select a reference. Press Next.
- Choose the reference files. Press Next.
- Optional Step - select the matching Known Variants vcf files for the reference files. Note that we need the name of the contig (e.g. chr17) to appear in the name of the matching .vcf file - here is an example of one we use within one of the Profiles: dbsnp-hg19-chr17.vcf. Press Next.
- Select Annotation Type - Exon/Gene/Amplicon. Press Next.
- Select a .bed or .gff file from the file system, containing the annotations that describe your Target.
- Select a snpEff database .bin file. Press Finish.

Reference Genome Dashboard

This is where the data for the Reference Genomes is managed.

In here, the actual fasta data for the reference genome (or chromosome) can be imported. Contigs will be automatically created when the fasta file is imported into the tool. We strongly recommend that whole chromosomes are used as the reference sequence, to allow for any off-target reads that might fall on pseudogenes to be properly mapped to the pseudogene and not cause false positives by being incorrectly mapped to a gene.

Known variants from dbSNP can also be imported, in .vcf format. These are important for the quality of the Variant Call, and it's recommended to import them.

It's also possible to import gene and exon annotations, in .bed or .gff format. These can be used for creating new Targets within the tool. If you don't need to create a new Target then you don't need to worry about this function.

Create Reference

This wizard is used to create a new Reference Genome within Omixon Target. Once created, this Reference will become available for all Experiments and Analyses to use.

The options in this wizard:

- Choose Reference Name
- Select Species

It is not sufficient to simply create a Reference Genome - before it can be used with in the tool it should also have reference data, known variants and gene/exon annotations imported.

References

Centralised reference configuration for the whole tool

Omixon Target has a centralised configuration for the References (reference sequences) used in the Data Analysis portion of the tool.

These References can be set up once, and then used in multiple Experiments and Analyses.

This screen lists the available Reference Genomes, and allows Reference Genomes to be created or deleted.

Clicking on a Reference Genome will take you to the Reference Genome Dashboard, where reference data (in fasta format) can be imported, along with known variants (from dbSNP http://www.ncbi.nlm.nih.gov/SNP/).

Target Dashboard

Targets can be configured within this dashboard.

Targets are optional within the tool, but are highly recommended, as there are a number of analysis features that give better results for well-defined targets.

A typical Target could be 'all the exons within the BRCA1 and BRCA2 genes'.

At the moment, the target configuration is restricted to gene, exon and amplicon annotations. There are two ways to import annotations - either via the 'Import Target' method, where annotations can be imported and a Target created in a single step, or the annotations to be used can be imported via the 'Import Annotations' function in the Reference Genome Dashboard - this second method requires a Target to be created or altered manually in order to make use of these annotations.

Targets

Centralised Target configuration for the whole tool.

Omixon Target has a centralised configuration for the Targets (essentially lists of annotations) used in the Data Analysis portion of the tool.

These Targets can be set up once, and then used in multiple Experiments and Analyses. Targets are actually optional, but recommended.

This screen lists all the Targets defined, and allows Targets to be created or deleted. It is possible to create a Target by importing annotations in BED or GFF format using the 'Import Target' function.

Clicking on a Target will take you to the Target Dashboard, where the Target can be configured.

Create Target

The wizard will allow the creation of a new Target. The other ways to create a target are to use the 'Import Target' wizard (and import annotations plus create a Target in a single step), or the 'Create Profile' wizard, which merges the 'Import Reference Data', 'Import Known Variants' and 'Import Target' wizards into one.

A Target is defined as a list of annotations. These annotations can be for genes, exons or amplicons. The annotations belong to a particular Reference Genome and can be imported into the tool via the Reference Genome Dashboard.

List of wizard steps:

- Create Target

It is not sufficient to simply create a Target - once a Target has been created it should be configured by clicking on it (which will navigate to the Target Dashboard), and then configuring the Target via the Configure Target wizard.

Configure Target

This function can be used to add and remove annotations from a Target. The annotations will already need to exist - either imported against a Reference, from the References screen, or by using the 'Import Target Annotations' function in the Targets screen, or the 'Create Profile' function in the main Data Analysis screen.

Experiment Dashboard

The Experiment is the top level of the small hierarchy organising the genomic data within the tool.

This is a flexible container, and should be used to reflect the nature of your analysis requirements.

An Experiment is a grouping of Analyses. Common configuration elements for a group of Analyses can be specified at the Experiment level, including the references to use, the sequencer, and the targets. This will restrict the Analysis configuration options, and later the options available to the various analysis tasks with the tool.

The Experiment can also be used for analysing and reporting the results of multiple Analyses together.

Design Experiment

This wizard usually only has to be run once, when a new Experiment is created.

This wizard inherits all its settings from the Profiles have been added to the tool, and the lists of configured Reference Genomes and Targets.

The steps in the wizard:

- General Properties
- Choose Sequencer
- Choose Target
- Choose Species
- Choose Genome
- Choose Contig

Once all these properties are set for an Experiment, there will be automatically used by Analyses that belong to the Experiment.

Experiment References

Configure References for an Experiment

Omixon Target has a centralised configuration for the References (reference sequences) used in the Data Analysis portion of the tool.

These References can be set up once, and then used in multiple Experiments and Analyses.

This screen lists the Reference Genomes linked to this Experiment, and allows Reference Genomes to be created or deleted.

Clicking on a Reference Genome will take you to the Reference Genome Dashboard, where reference data (in fasta format) can be imported, along with known variants (from dbSNP http://www.ncbi.nlm.nih.gov/SNP/).

Experiment Targets

Configure Targets for an Experiment

Omixon Target has a centralised configuration for the Targets (essentially lists of annotations) used in the Data Analysis portion of the tool.

These Targets can be set up once, and then used in multiple Experiments and Analyses.

This screen lists the Targets attached to this Experiment, and allows Targets to be created or deleted. It is possible to create a Target by importing annotations in BED or GFF format using the 'Import Target' function.

Clicking on a Target will take you to the Target Dashboard, where the Target can be configured.

Alignment and Variant Call

Analysis Dashboard

The main purpose behind the Analysis is to group together a set of samples.

These Samples will share common configuration within the tool - they will share the same targets, the same references, the same sequencer etc.

The Analysis inherits its configuration from its parent Experiment, and can be used to help split the Experiment into more specific sections. There could be one Analysis for Ion Torrent and one for Illumina data, for example, or one Analysis for one Target (e.g the HLA-A gene) and another for another Target (e.g. the HLA-B gene).

From the Analysis Dashboard, it is possible to browse through and work with individual Samples. It's also possible to select multiple Samples and work with these together, either by launching them in the Genome Browser, or by comparing them using the 'Compare Samples' button, which will launch the Sample Difference screen.

Design Analysis

This wizard usually only has to be run once, when a new Analysis is created.

This wizard inherits all its settings from the parent Experiment, and can be used to restrict the options from the Experiment for each Analysis.

The steps in the wizard:

- General Properties
- Choose Sequencer
- Choose Target
- Choose Species
- Choose Genome
- Choose Contig

Once all these properties are set for an Analysis, there will be automatically used by all the wizards and processes for the Samples that belong to the Analysis.

Analysis References

Configure References for an Analysis

Omixon Target has a centralised configuration for the References (reference sequences) used in the Data Analysis portion of the tool.

These References can be set up once, and then used in multiple Experiments and Analyses.

This screen lists the Reference Genomes linked to this Analysis, and allows Reference Genomes to be created or deleted.

Clicking on a Reference Genome will take you to the Reference Genome Dashboard, where reference data (in fasta format) can be imported, along with known variants (from dbSNP http://www.ncbi.nlm.nih.gov/SNP/).

Analysis Targets

Configure Targets for an Analysis

Omixon Target has a centralised configuration for the Targets (essentially lists of annotations) used in the Data Analysis portion of the tool.

These Targets can be set up once, and then used in multiple Experiments and Analyses.

This screen lists the Targets attached to this Analysis, and allows Targets to be created or deleted. It is possible to create a Target by importing annotations in BED or GFF format using the 'Import Target' function.

Clicking on a Target will take you to the Target Dashboard, where the Target can be configured.

Map and Align Samples

Introduction

If you have imported sequencing data, then you can map and align this data to a Reference Genome using this function.

This version of the Map and Align runs with multiple samples - i.e. all the samples selected in the list visible in the Analysis Dashboard. If any of the samples selected already have mapped data and/or variant calls they will be (silently) ignored.

This function has three analyses built in to it:

- Map and Align using Omixon mapper or BWA
- Variant Call using GATK or Samtools (optional)
- Variant Effect Prediction using snpEff (optional)

The Map and Align step will always be performed. The Variant Call and Variant Effect Prediction steps are optional. The Variant Call and Variant Effect Prediction steps can be performed later using the 'Call Variants' function, or the mapped data can be copied elsewhere and other variant caller used instead.

The steps in this wizard:

- Select Reference Data
- Select Sequencer
- Advanced Options
- Variant Call Options
- Variant Effect Options

Like all the wizards within the tool, most of the options within this wizard will already be pre-filled and pre-selected, based on your chosen Analysis and Experiment configuration.

Map and Align

Alignments can be run using Omixon's own aligner algorithm (ORM) or - in the Pro version - the BWA short read or long read algorithms.

- Parameters file: You can choose to import a properties file with some advanced parameters. This file is the same one used for the Omixon Variant Toolkit (the command line version of our alignment algorithms), you can find the latest readme file for this linked via the Omixon web site (the readme file link is in the 'Useful Links' section at the bottom of the page): https://www.omixon.com/omixon/abouttoolkit.htm
- Maximum coverage: You can choose what the maximum depth of coverage should be for your results. For very deep coverage data, it's usually enough to keep 1000 deep short reads, however if you want to you can keep more or less than this. This will cause your actual mapping results to be discarded, only a number matching your maximum coverage will be kept.
- Low coverage threshold: the threshold used for low coverage annotations in the "Coverage annotation" function.
- High coverage: the threshold used for high coverage annotations in the "Coverage annotation" function.

Note, that coverage annotation thresholds can be changed at any time on the "Analyse sample" page, using the "Coverage annotations" function. When the thresholds are changed, the annotations are automatically recalibrated using the new values.

Variant Call

You can choose whether or not to run a GATK variant call after the map and align function has run. You can always run the variant call later, if desired.

- The dbSNP (known variants) file is an optional parameter. You can either use dbSNP data that has been imported into the tool, or select a dbSNP file from the file system.

In the Pro version of Omixon Target, a Samtools variant call is also available from this wizard. Samtools is the suggested variant caller for long reads (Ion Torrent and 454).

Variant Effect Prediction

In the Pro version, variant effect prediction can be run within Omixon Target. Variant annotations are generated using snpEff (SNP Effect Predictor). For HG19 chromosomes, a variant effect database is provided, for other references a custom database can be selected from the file system. The most common codon sets are provided and can be selected in the wizard. Custom codon sets can be specified as well in an external file. Advanced parameters can be chosen using a text file.

Sample Difference

This is the starting point for multiple sample comparison, for example with trio data (mother, father, child).

This screen is started by selecting multiple Samples within the Analysis Dashboard, and then selecting 'Compare Samples'.

Variants within the comparison table can be selected, and then you can 'browse' that position within the Genome Browser, which will also allow you to examine all the samples being viewed in the Samples Difference screen together.

Analysis Results

Summary statistics for all the Samples in the Analysis

This is a very simple summary screen that is available from the Analysis Results button in the Analysis Dashboard. It displays overview statistics of the progress for all the Samples in the Analysis.

Sample Dashboard

A Sample represents a single set of sequencing data for a single individual.

The Sample has a small life-cycle within the tool:

- Create Sample
- Identify Variants
- Approve Variants
- Approve Sample
- Analyse Approved Sample and Approved Variants

The concept is that the 'raw' variants called may not all be of sufficiently high quality for downstream analysis, and the tool supports a number of Approval functions in order to approve or reject variations and filter them down a set suitable for further analysis. In the case of a HLA Typing analysis, then the downstream analysis step will be to determine the HLA Types of the sample.

Create Sample

This wizard allows you to create an empty Sample. You need to give it a name (which should be unique within the Analysis) and select a species. Once a Sample is created, data can be imported for analysis.

The other way to create Samples is to use the 'Import multiple samples' function from the Analysis Dashboard, which will both import fastq and automatically create Samples at the same time.

Map and Align

Introduction

If you have imported sequencing data, then you can map and align this data against a Reference Genome using this function. This version of the Map and Align runs with a single sample. This function has four analyses built in to it:

- Map and Align using Omixon mappers
- Map and Align using Burrows-Wheeler aligner (only in Target Pro edition)
- Variant Call using GATK (optional) - preferred for Illumina data
- Variant Call using Samtools (only in Target Pro edition) - preferred for Ion Torrenet and Roche 464 data.
- Variant Effect Prediction using SnpEff (only in Target Pro edition).

The Map and Align step will always be performed. The Variant Call and the Variant Effect Prediction steps are optional. Both of these can be performed later using the 'Call Variants' function, or the mapped data can be copied elsewhere and other variant caller used instead.

The steps in this wizard:

- Select Genomic Data
- Select Sequencer
- Advanced Options
- Variant Call Options
- Variant Effect Options

Like all the wizards within the tool, most of the options within this wizard will already be pre-filled and pre-selected, based on your chosen Analysis and Experiment configuration.

Map and Align

Alignments can be run using Omixon's own aligner algorithm (ORM) or - in the Pro version - the BWA short read or long read algorithms.

- Parameters file: You can choose to import a properties file with some advanced parameters. This file is the same one used for the Omixon Variant Toolkit (the command line version of our alignment algorithms), you can find the latest readme file for this linked via the Omixon web site (the readme file link is in the 'Useful Links' section at the bottom of the page): https://www.omixon.com/omixon/abouttoolkit.htm
- Maximum Coverage: You can choose what the maximum depth of coverage should be for your results. For very deep coverage data, it's usually enough to keep 1000 deep short reads, however if you want to you can keep more or less than this. This will cause your actual mapping results to be discarded, only a number matching your maximum coverage will be kept.
- Low coverage threshold: the threshold used for low coverage annotations in the "Coverage annotation" function.
- High coverage: the threshold used for high coverage annotations in the "Coverage annotation" function.

Note, that coverage annotation thresholds can be changed at any time on the "Analyse sample" page, using the "Coverage annotations" function. When the thresholds are changed, the annotations are automatically recalibrated using the new values.

Variant Call

You can choose whether or not to run a GATK variant call after the map and align function has run. You can always run the variant call later, if desired.

- The dbSNP (known variants) file is an optional parameter. You can either use dbSNP data that has been imported into the tool, or select a dbSNP file from the file system.

In the OmixonTarget Pro edition, you can also choose to run a Samtools mpileup variant call instead of a GATK variant call. This is recommended for 454 and IonTorrent data.

Variant Effect Prediction

In the Pro version, variant effect prediction can be run within OmixonTarget. Variant annotations are generated using snpEff (SNP Effect Predictor). For HG19 chromosomes, a variant effect database is provided, for other references a custom database can be selected from the file system. The most common codon sets are provided and can be selected in the wizard. Custom codon sets can be specified as well in an external file. Advanced parameters can be chosen using a text file.

Call Variants

Introduction

If the Map and Align function is used to map short reads, then there is an option to run a Variant Call within that wizard. If that option is not chosen, or if already mapped reads are imported into the tool, then a separate Variant Call can be started.

The Variant Call invokes a full GATK pipeline, following the recommended best practises of the Broad Institute. In the Pro version of OmixonTarget, a Samtools mpileup variant call can be selected as well. This is the preferred variant calling option for Ion Torrent and Roche 454 data.

Like all wizards in the tool, as many options as possible are already preselected within the wizard.

These are the wizard steps:

- Select Genomic Data
- Select Sequencer
- Advanced Options
- Variant Call Options
- Variant Effect Options

Advanced Options

Parameters file: This is currently not used for the variant caller.

Variant Call Options

The dbSNP (known variants) file is an optional parameter. You can either use dbSNP data that has been imported into the tool, or select a dbSNP file from the file system.

Variant Effect Prediction

In the Pro version, variant effect prediction can be run within Omixon Target. Variant annotations are generated using snpEff (SNP Effect Predictor). For HG19 chromosomes, a variant effect database is provided, for other references a custom database can be selected from the file system. The most common codon sets are provided and can be selected in the wizard. Custom codon sets can be specified as well in an external file. Advanced parameters can be chosen using a text file.

Analyse Sample

Introduction

This screen is the starting point for the analysis of the variants found in a Sample.

If there are known ('expected') variants already found for this sample, these can be compared with the actual variants found during the analysis.

The variants found are listed according to whether they are on or off target, and it's possible to select individual variants and jump into the Genome Browser to inspect them.

It's possible to start the Approval Dashboard from here, in order to approve the variants ready for further analysis (and reject any low

quality or off target variants that should be ignored).

It's also possible to navigate to the Report Dashboard to see graphs, charts and other summary information about the Sample data (currently under development).

The Coverage report can also be reached from this screen. This report contains statistics about the coverage of the sample. A coverage based annotation tool is also available from here.

Summary

On the "Summary" tab, a short list of basic information about the selected sample is shown:

- Name of the file that contains the variant call results.
- Name of the file that contains the expected variants.
- Name of the currently selected contig. (Tip, you can chose a different contig using the "Switch contig" button.)
- Total number of variants found on the current contig.
- Total number of variants found on all the contigs.

On-target Variants

On this tab, all variants located in the specified target are listed (you can check the targets on the "Analysis targets" screens). For each variant, the following fields are always available:

- Start pos: the start position of the variant.
- End pos: the end position of the variant.
- Ref value: the reference bases at the location of the variants.
- Type: the type of the variant (SNP or INDEL).
- Value: alternative value (i.e. the nucleotide(s) supported by the reads).
- Frequency: frequency of the alternative allele in the reads.
- Quality: quality of the variant.
- Coverage: coverage at the position of the variant.
- Id: dbSNP ID of known variants.
- Status: status of the variant, possible values: "Pending", "Approved" and "Rejected".

Note, that when the variant effect prediction function is run, additional fields are shown between "Coverage" and "Id". For a description of these additional fields, see the "Variant effect results" section of this page.

Off-target Variants

On this tab, all variants located outside the specified target are listed (you can check the targets on the "Analysis targets" screens). For a short description of the fields shown in the table, see the "On-target Variants" section.

Variant effect results

Variant effect prediction results are generated using snpEff (Pro Server version only). These results are shown on the "Analyse sample" page, in the "On-target variants" and "Off-target variants" tables. The snpEff result fields are the following:

- Effect: effect of the variant.
- Impact: estimated impact of the variant (possible values are "Modifier", "Low", "Moderate" or "High"). The impact is estimated based on the type of the effect. For details about impact categories, see the table below.
- Region: the coding region where the variant is located in (according to the SnpEff database).
- Codon: the codon change caused by the variant (reference codon/alternative codon, the variant is shown in upper case, the rest of the codon is shown in lower case).
- Aa: amino-acid change. (One letter code of the reference amino-acid, followed by the location of the amino-acid, followed by the one letter code of the alternative amino-acid. If the mutation is synonymous, the second letter is missing.)

The original snpEff results are shortened, to fit into the variant call table. For a full list of the abbreviations used, see the list below.

Original result - Omixon Target Result - Estimated impact

- INTERGENIC - Intergenic - Modifier
- UPSTREAM - Upstream - Modifier
- UTR_5_PRIME - Utr5 prime - Modifier
- UTR_5_DELETED - Utr5 deleted - Modifier
- START_GAINED - Start gained - Low
- SPLICE_SITE_ACCEPTOR - Splice acceptor - High
- SPLICE_SITE_DONOR - Splice donor - High
- START_LOST - Start lost - High
- SYNONYMOUS_START - Syn start - Low
- CDS - Cds - Modifier
- GENE - Gene - Modifier
- TRANSCRIPT - Transcript - Modifier
- EXON - Exon - Modifier
- EXON_DELETED - Exon deleted - High
- NON_SYNONYMOUS_CODING - Non syn coding - Moderate
- SYNONYMOUS_CODING - Syn coding - Low
- FRAME_SHIFT - Frame shift - High

- CODON_CHANGE - Codon change - Moderate
- CODON_INSERTION - Codon insert - Moderate
- CODON_CHANGE_PLUS_CODON_INSERTION - Codon ch+ins - Moderate
- CODON_DELETION - Codon delete - Moderate
- CODON_CHANGE_PLUS_CODON_DELETION - Codon ch+del - Moderate
- STOP_GAINED - Stop gained - High
- SYNONYMOUS_STOP - Syn stop - Low
- STOP_LOST - Stop lost - High
- INTRON - Intron - Modifier
- UTR_3_PRIME - Utr3 prime - Moderate
- UTR_3_DELETED - Utr3 deleted - Moderate
- DOWNSTREAM - Downstream - Modifier
- INTRON_CONSERVED - Intron cons - Modifier
- INTERGENIC_CONSERVED - Inter cons - Modifier
- INTRAGENIC - Intragenic - Modifier
- RARE_AMINO_ACID - Rare amino acid - High
- NON_SYNONYMOUS_START - Non syn start - Low

Coverage Annotations

The coverage report contains some custom annotations, for target regions where the coverage is too high, too low or zero.

These can be regenerated using different threshold values for the high and low coverage limits - i.e. target regions with higher coverage than the 'High coverage threshold' and target regions with lower coverage than the 'Low coverage threshold' will have new annotations created. Any existing high/low coverage annotations will be discarded.

Approval Dashboard

Not all variants are interesting. Not all variants can be trusted.

Actual variants that have been discovered during the Variant Call can vary in quality. Some of them may be off-target and therefore not interesting. Some may only have very low coverage and so are not trustworthy. Some may be complex and require manual inspection to figure out what is actually going on in the underlying short read data.

Variants can be 'approved' or 'rejected' in the Approval Dashboard. All variants start in 'pending' status, and there are functions to automatically approve and reject groups of variants by various criteria, and also to manually approve or reject individual variants as well.

Downstream analysis steps will only consider approved variants for their input.

Auto Approve

Even with targeted sequencing, there can be a large number of variants to deal with.

Auto approve allows the 'bulk' approval of variants that meet one or more of a number of criteria, including:

- The quality of the call is above a minimum quality
- The coverage at the location is greater than a minimum coverage
- The call is on Target

Auto Reject

Even with targeted sequencing, there can be a large number of variants to deal with.

Auto reject allows the 'bulk' rejection of variants that **fail** to meet one or more of a number of criteria, including:

- The quality of the call is above a minimum quality
- The coverage at the location is greater than a minimum coverage
- The call is on Target

Auto Reset

If you have made a mistake with one of the Auto Accept or Auto Reject functions, you can reset all your variants back to 'pending' status by using the Auto Reset function. This is an all-or-nothing operation.

Data Analysis Module Browser

Genome Browser

Introduction

Multiple Samples can be browsed together.

The data items that can be visualised include:

- Reference Sequence
- Gene Annotations
- Exon Annotations
- Amplicon Annotations

And for each Sample:

- Actual Variant Annotations
- Expected Variant Annotations
- Short Reads, including Coverage
- Low/High/Zero coverage annotations

By default, the display is 'masked' so that only differences between the short reads and the reference sequence are displayed. This mask is done in a stranded fashion, so that forward strand reads (pink) and reverse strand reads (yellow) can be easily distinguished within the display. The 'Display Strand' function turns this on and off.

The short reads track can be 'collapsed' which gives a summary view of the short reads (and does not allow each read to be inspected in detail).

Individual items (annotations, short reads) within the Genome Browser can be selected, and it's possible to browse elsewhere and then jump back to the selected item, or copy the details of the selected item to the clipboard.

Settings and Functions

Expanding the display

At the top left of the tool is a small control that allows an expanded view. This actually works for all screens in the tool, but is most useful for the Genome Browser. Clicking the same icon again will go back to the usual display.

Rotating the Genome Browser

The vertical Genome Browser view is more useful for comparing multiple samples. The view can be rotated to use a horizontal display, which is a bit more traditional for Genome Browsers, and is more useful when browsing a single sample.

Filtering the Short Reads

You can display more short reads by using the 'collapsed' pile-up view.

You can also 'page' around the short reads using the small arrows at the top of the short read track. You can set the page size using the up and down arrows, and the current page by using the left and right arrows. This is useful when you have deep sequencing and want to scroll through a few hundred short reads at a time.

Display setup

The variants in the 'Act' (Actual) variants track can also be filtered. This makes use of the 'Approval' features within the tool - individual variants can be 'approved' or 'rejected' (or still be 'pending', awaiting either approval or rejection). This filter will display 'All' variants (default), then 'Pending', 'Approved' or 'Rejected' variants. The visualisation of variants in the short read track can be customized as well. Variants are always visible on the nucleotide level, but insert and deletion marks can be selectively turned on and off. Soft clips can be shown or hidden as well.

Exporting Genome Browser Data

You can 'Copy to Clipboard' the details of individual, selected items in the display. You can select a read and by clicking on the 'Copy to clipboard' button (which can be found on the middle panel in the bottom of your screen), you can copy the name and CIGAR string of the selected read to the clipboard. Similarly, for a selected annotation, the position and name of the annotation can be copied to the clipboard.

You can also copy the reference sequence belonging to a selected item (read or annotation), by clicking on the 'Copy reference to clipboard' button.

You can also 'Capture a Screenshot' of the currently visibly tracks within the Genome Browser. This will create a .png file at your chosen location.

Working with Tracks

Tracks used in the Genome Browser

There are two different kinds of tracks used within the Genome Browser. **"Built-in tracks"** can be created using different parts of the Data Analysis module (e.g. importing a reference file, or running an alignment or variant call). These tracks are not editable and are automatically added to the Omixon Target database. The other kind of tracks are called **"User tracks"**. These tracks can be editable and are stored in external (non database) files.

Creating tracks

User tracks can be created within Omixon Target (using the 'Create Track' button in the 'Manage Tracks' wizard), or can be imported from external files (using the 'Add Track' function in the 'Manage Tracks' wizard). User tracks created within Omixon Target are always editable. User tracks imported from file can be editable or non-editable (you can set this when importing the file). Annotation editing is currently only supported for files in bed format.

Modifying tracks

Currently, track editing is only available for bed formatted, editable, user defined tracks.

- ***To create a new annotation entry:*** use the green '+' button on the middle part of the bottom panel, or simply press the 'Insert' button on your keyboard. This will automatically generate an annotation entry based on the currently shown browser region. You can choose which track you want to add this newly defined entry to, from a list containing all the editable tracks and you can also modify the name and end coordinates of the annotation entry by hand.
- ***To modify an existing annotation entry "by hand":*** you have to select it first (this can be done simply by clicking on the entry with the left mouse button). The selected annotation element is marked by a lighter colour. The attributes of the annotation element can be edited using the 'Edit selected annotation' function (this button is located on the middle bottom panel, next to the 'Create new annotation' button). You can also open the annotation editor screen by pressing the ENTER button, after selecting an annotation.
- ***To resize an annotation entry in the browser*****:** select the annotation, lock the cursor (you can do this by pressing 'Space' on your keyboard), then press the 'Resize selected annotation' button on the middle bottom panel or press 'R' on the keyboard. The selected annotation will be resized, to end at the cursor position.
- ***To move an annotation entry in the browser:*** select the annotation, lock the cursor (you can do this by pressing 'Space' on your keyboard), the press the 'Move selected annotation' button on the middle bottom panel or press 'M' on the keyboard. The selected annotation will be moved to the cursor position.

Deleting tracks

To delete an annotation entry, select the annotation, then press the 'Delete selected annotation' button (red 'X' symbol on the middle bottom panel) or the 'Delete' button on your keyboard.

Showing or hiding tracks

All tracks (both built-in and user created tracks) can be temporarily hidden using the 'Hide track(s)' button in the 'Track Manager' wizard.

Navigation in the Browser

Zooming

By default the Genome Browser starts in 'Drag' mode, where the mouse wheel (or '+' and '-' keys) can be used to zoom in and out, and the display can be moved by clicking and dragging with the mouse. 'Drill' mode can also be used, where the region highlighted by a mouse click and drag will be 'drilled into' once the mouse button is released.

Jump to a position

You can jump to a position in the reference using 'Jump To Position', or jump to one of the annotations within the Target using 'Jump To Target' or 'Jump to annotation'. If you have previously selected a feature within the display (such as an annotation or short read), then you can jump back to that feature at any time by using the small arrow button at the bottom of the display.

The 'Jump To Position' wizard can be opened by pressing the "CTRL" and "J" keys on the keyboard. After the wizard is opened, you can type in a single coordinate or a pair of coordinates and you can specify a reference contig. When a reference contig is not chosen, the currently shown contig is used. Coordinates can be specified with or without thousand delimiters (i.e. 15200, 15.200 and 15,200 are all accepted and mean the same).

Some examples (the currently selected reference contig is chr1):

- ***CTRL+J 15200 ENTER*** -> The Genome Browser display will jump to chr1:15200.
- ***CTRL+J 15.200-15.400 ENTER*** -> The Genome Browser display will show the area chr1:15200 - 15400 wizard can be opened by pressing the "CTRL" and "J" keys on the keyboard. After the wizard is opened, you can type in a single coordinate or a pair of coordinates and you can specify a reference contig. When a reference contig is not chosen, the currently shown contig is used. Coordinates can be specified with or without thousand delimiters (i.e. 15200, 15.200 and 15,200 are all accepted and mean the same thing).
- ***CTRL+J chr2:15,200 ENTER*** -> The Genome Browser display will jump to chr2:15200.

Jump to an annotation

The 'Jump To Annotation' wizard can be opened by pressing the "CTRL" and "F" keys on the keyboard. After the wizard is opened, you can type in the name (or the beginning of the name) of an annotation (e.g. gene, exon, amplicon). When the specified search term is ambiguous (i.e. more than one annotation name starts with the specified string), a list of annotations is presented and the desired annotation can be selected from the list. The input field in this wizard is not case sensitive (i.e. brca1 and BRCA1 are both accepted and mean the same thing).

Search for a nucleotide sequence

By pressing CTRL+S, you can open the 'Search sequence' wizard. Here you can specify a nucleotide sequence to search for in the currently visible reference contig and then step through the hit positions.

Shortcuts

Keyboard/mouse navigation options and shortcuts

- **UP/DOWN arrow keys**: scroll alongside the reference contig.
- **LEFT/RIGHT arrow keys**: scroll through the reads at the current position/area.
- **PAGE UP/PAGE DOWN**: scroll alongside the reference contig in bigger steps.

- **HOME/END**: jump to the beginning/end of the current reference contig.
- **+/-**: zoom in and out.
- **SPACE**: locks/unlocks the cursor (i.e. green line when unlocked, red line when locked).
- **BACKSPACE**: shows the whole reference contig (i.e. zooms out totally) or when the cursor is locked, jumps to the locked cursor.
- **Mousewheel**: zoom in/out (in 'Drag' mode) or move alongside the reference (in 'Drill' mode).
- **TAB/SHIFT+TAB**: jump to next/previous reference contig.
- **SHIFT+mousewheel**: scroll through the reads at the current position/area.
- **Left mouse button**: can be used for dragging the display (in 'Drag' mode) or for highlighting and zooming in on a region (in 'Drill' mode by clicking and releasing the left button).
- **CTRL+J**: opens the 'Jump to position' wizard. Here you can specify a reference contig and a coordinate or area on that contig and jump to the selected location.
- **CTRL+F**: opens the 'Jump to annotation' wizard. Here you can specify the name (or partial name) of an annotation (e.g. gene, amplicon, exon) and jump to the selected annotation.
- **CTRL+S** : opens the 'Search sequence' wizard. Here you can specify a nucleotide sequence to search for in the currently visible reference contig and then step through the hit positions.
- **CTRL+M**: enter full screen mode - hide all menus and show the browser region in full screen.
- **ESCAPE**: exit full screen mode.
- **X**: switches the browser between horizontal and vertical mode.
- **C**: switches the browser between collapsed and uncollapsed mode.
- **T**: switches the browser between drag and drill navigation mode.
- **INSERT**: insert new annotation.
- **ENTER**: edit selected annotation.
- **DELETE**: delete selected annotation.
- **M**: move selected annotation to start at the locked cursor position.
- **R**: resize selected annotation to end at the locked cursor position.

Manage Tracks

In this wizard, you can hide and show any tracks, and import, create and delete additional annotation and variant tracks.

The visibility of any track can be set with the 'Show Track(s)' and 'Hide Track(s)' functions. Hidden tracks are marked with a darker grey color in the 'Manage Tracks' wizard track list.

Annotation tracks created by hand or added from file in this wizard are called **"User tracks"**. These tracks can be edited within the Genome Browser.

Empty tracks can be created using the 'Create Track' button. These tracks are always editable.

Predefined (i.e. non-empty) tracks can be imported from file using the 'Add Track' button. Annotations must be in gff or bed format, variant calls are accepted in vcf format. The editability of these tracks can be set during the import process. For non-empty tracks, the type of the imported annotations (variant, exon, gene or amplicon) can be set as well. The default annotation type is "exon" for gff and bed files and "variant" for vcf files.

Create Track

In this wizard, an empty annotation track can be created within Omixon Target. The name and type of the annotation can be set (the annotation type can be exon, gene or amplicon). This user defined annotation track is stored in a bed formatted file, at a location selected by the user. When the annotation track is edited within the Genome Browser, this storing bed file is automatically updated and saved.

Add Track

In this wizard, annotation tracks (in vcf, gff or bed format) can be imported from external files. The type of the imported track (variant, exon, gene or amplicon) and the 'Fallback track name' (i.e. the name shown in the Browser track header) can be specified. For vcf files, the track type is automatically set to 'Variant annotation'. The editability of the imported track can be set as well. (Note, that editing is only supported for bed formatted files at the moment.)

Import and Export

Import Reference Data

This function will import the reference data for the Reference Genome.

It is recommended to import and use whole chromosomes. Even targeted sequencing can place some reads off target, and it's very important that off target reads are also mapped off target correctly, particularly if they happen to map to a pseudogene of the gene(s) being targeted.

The input file must be in fasta format. We recommend using multiple fasta files within Omixon Target, rather than Multifasta references.

Contigs will be automatically created, based on the names found within the input fasta file.

Import Known Variants

This function will import dbSNP known variant annotations for the Reference Genome.

This will be used later with the GATK variant call.

If the Reference Genome has been configured to use a whole chromosome, then we recommend importing all matching dbSNP records for that chromosome.

The annotations must be in .vcf format.

Import Variant Effect

In this wizard, an SnpEff variant effect prediction database can be added for a reference. The database must be in the format specified by the SnpEff documentation. If a variant effect database file is added, the entries required for the SnpEff run are automatically created in the snpEff config file. In any built-in profiles are added, or HG19-based profiles are created, the HG19 SnpEff database is automatically downloaded and added to the reference.

Tip: Custom databases can be created for any kind of reference sequence. For details, take a look at the SnpEff documentation.

Import Annotations

This function will import Gene or Exon annotations for the Reference Genome.

This will be used later for Target configuration.

Currently .bed and .gff formats are supported. If you need another format please let us know.

The steps in this wizard are:

- Select Annotation Type
- Select Input Files
- Select Contig

Import Target Annotations

This function will import annotations for genes, exons or amplicons and use them to create a Target.

The reference genome will need to be created first (if it doesn't exist already), and all the reference data should already be imported within Omixon Target.

The easier alternative to this function is to use the 'Create Profile' wizard, which merges the 'Import Reference Data', 'Import Known Variants' and this 'Import Target' wizard into one function.

Import Sequencing Data

The first goal of Sample analysis is to identify some variants. There are three ways to do this, and this function allows one of these three ways.

If you have short read data from sequencing run you can import it with this function, and then run the 'Map and Align' function, which will align the data to a Reference Genome and run a variant call in order to identify the variants.

You will usually either run this wizard from within an existing Sample, or use the Import Multiple Samples wizard from the Analysis Dashboard, which both imports data and creates (multiple) samples at the same time.

This function will allow you to import two FASTQ files at the same time (or one SFF). The second one is optional, and is for paired reads if your analysis produced paired end or mate pair data. If using paired reads, the two input files are assumed to have the exact matching reads, in the exact same order (there is currently no in-built check for this). The SFF import doesn't support paired mode and is only available in Omixon Target Pro.

Steps in this wizard:

- Select Input File
- Select Paired Read Options

After you have imported sequencing data, you can align it and call variant in the data using the Map and Align function.

Paired Read Options

- Minimum Distance - the minimum expected distance between the two reads
- Maximum Distance - the maximum expected distance between the two reads
- Orientation - the orientation of the reads with respect to each other. The first file is the first in the orientation as well. Options are FR (forward-reverse, i.e. first file contains forward reads, second file contains reverse reads), RF (reverse-forward) and FF (forward-forward).

Import Mapped Data

The first goal of Sample analysis is to identify some variants. There are three ways to do this, and this function allows one of these three

ways.

If you have short read data that you have already mapped from another source you can skip the 'Import Sequencing Data' and 'Map and Align' functions altogether, and just import your mapped data into the tool (in either .sam or .bam format), and then use the Call Variants function to identify the variants.

Steps in this wizard:

- Select Input File
- Select Reference Contigs
- Select a Target Filter - this will cause only data that overlaps your target region(s) to be imported
- Select maximum coverage - i.e. choose how much coverage to import

The maximum coverage feature should be used with care, it will take the first X reads/pairs it finds, it's not a sampling method.

After you have imported mapped data, you can browse it via the Genome Browser, or use the Call Variants function to identify variants within the mapped data.

Import Variant Calls

The first goal of Sample analysis is to identify some variants. There are three ways to do this, and this function allows one of these three ways.

If you have already mapped your data and called variants on it, or if you have variants already identified from another source (such as Sanger sequencing), then you can skip the 'Map and Align' and 'Call Variants' functions altogether and simply import your variants into the tool.

This function also allows you to import 'expected' variants. If you are working with simulated data, or if you already have known variants from another source then you can import these and tag them as 'expected' variants. When you have then identified variants (via an import of 'actual' variants, or via a Map and Align or Call Variants analysis), then you can automatically compare the expected variants with the actual variants using the Analyse Sample Variants screen.

Steps in this wizard:

- Select Annotation Type
- Select Input Files
- Import Options
- Select Contig

After you have imported variant calls, you can visualise them in the Genome Browser, analyse them within the Analyse Sample Variants screen, and compare them with the variant calls of other Samples using the Compare Samples function.

Import Multiple Samples

This function is available from the Analysis Dashboard.

If you have short read data from sequencing run you can import it with this function, and then run the 'Map and Align' function, which will align the data to a Reference Genome and run a variant call in order to identify the variants.

It is essentially identical to the 'Import Sequencing Data' function within the Sample Dashboard - the goal it to import single or multiplexed fastq/sff data into the tool. The difference with this version is that it will do two things:

- Import fastq/sff data (or paired fastq data)
- Automatically create a new Sample for each imported fastq file (or pair of fastq files) in case of single sample data
- Demultiplex each fastq/sff file and create a new Sample for each sample found in the files identified by barcodes

This function will allow you to import multiple fastq/sff files - or multiple pairs of matching fastq files - at the same time. The second file is optional, and is for paired reads - i.e. if your analysis produced paired end or mate pair data. If using paired reads, the two input files are assumed to have the exact matching reads, in the exact same order (there is currently no in-built check for this). The paired read import is not supported for sff files. Note that the sff import function is only available in OmixonTarget Pro edition.

Steps in this wizard:

- Select Input File
- Select Paired Read Options
- Demultiplexing Options (Omixon Target Pro edition only)

After you have imported sequencing data, you can align it and call variant in the data using the Map and Align function.

Paired Read Options

- Minimum Distance - the minimum expected distance between the two reads
- Maximum Distance - the maximum expected distance between the two reads
- Orientation - the orientation of the reads with respect to each other. The first file is the first in the orientation as well. Options are FR (forward-reverse, i.e. first file contains forward reads, second file contains reverse reads), RF (reverse-forward) and FF (forward-forward).

Demultiplexing Options (Omixon Target Pro edition only)

- Sequencer - the sequencer used to produce the multiplexed data
- Barcode file - text file containing barcode names and related sequences to identify the samples in the multiplexed files

Note that Illumina demultiplexing is supported only for paired fastq files and Roche 454 demultiplexing is supported only for not paired sff files. Ion Torrent data demultiplexing is not supported currently.

Import Multiple Mapped

This function is available from the Analysis Dashboard.

If you have already mapped some short read data you can import it with this function, and then run a variant call in order to identify the variants.

It is essentially identical to the 'Import Mapped Data' function within the Sample Dashboard - the goal it to import single sam/bam file into the tool. The difference with this version is that it will do two things:

- Import sam/bam data
- Automatically create a new Sample for each imported sam/bam file

This function will allow you to import multiple sam/bam files at the same time.

Steps in this wizard:

- Select Input File
- Select Reference Contigs
- Select a Target Filter - this will cause only data that overlaps your target region(s) to be imported
- Select maximum coverage - i.e. choose how much coverage to import

The maximum coverage feature should be used with care, it will take the first X reads/pairs it finds, it's not a random sampling method.

After you have imported mapped data, you can browse it via the Genome Browser, or use the Call Variants function to identify variants within the mapped data.

Export Mapped Data

Any mapped short read data that has been mapped within Omixon Target (or imported using the 'Import Mapped Data' function) can be exported again, either in .sam or .bam format.

The export will also create a standard index file for the exported data.

Export Mapped Regions

From the Genome Browser, it is possible export the short read that overlap or are contained within the currently visible region.

There are two options for the export - the reads can be exported in either .fastq format (for re-mapping) or in .sam/.bam format.

Export Approved Variants

This function is currently the final step in the 'data analysis' pipeline. Once you have a short-list of approved variants you can export an annotated .vcf file from the tool using this function. Only variants that have been 'approved' (either by the manual or auto approval functions) can be exported in this way.

HLA

HLA Typing Dashboard

All the HLA Typing functions are available from here.

Firstly, the 'Setup Typing' function should be run. This only needs to be run once, however this step can be re-run at any time and will update the HLA database if a new version is available.

Once the Setup Typing function has been run, the HLA Typing function becomes available. From here, **multiple samples can be typed together**, within a single 'analysis'.

It's possible to view the results, as well as to delete results from previous runs. **Results from multiple analyses can be viewed simultaneously**. (Tip - You can use CTRL or SHIFT clicks to select multiple analyses).

There are two display modes for the analyses, the default 'card' mode and searchable/sortable list mode. In the card display mode you can view the protocol (set of parameters) from any analysis run, and save this protocol with a new name.

In the header section of the dashboard, you can see the version and date of the IMGT/HLA database that is currently configured together with the sequencers which have been set up within the tool.

HLA Typing Setup

This only needs to be run once (unless you would like to use another sequencer).

- Choose the sequencer(s) you would like to use, and hit finish. (Tip - you can use CTRL and SHIFT clicks to select multiple sequencers.)

A configuration file will be downloaded from the internet and installed into Omixon Target. This should only take a few seconds.

Once this has finished you can start to use the HLA Typing function.

Download problems

If for some reason you cannot download the HLA configuration file from within the tool (e.g. due to network security settings) you can download the files manually and use the 'Select Local file(s)' option instead of the 'Download file(s)' option in the wizard.

The in-built HLA configuration file can be downloaded from here:

- http://omixon-download.s3.amazonaws.com/target_haplotype_db.zip

HLA Typing

Introduction

This function is available from the HLA Typing Dashboard. Before running HLA Typing, the Setup Typing wizard will need to be run.

You need to fill in a name for this HLA typing Analysis, plus select files for the input. You also need to select which sequencer was used to created the fastq data.

This function will allow you to **process multiple samples at the same time**, i.e. multiple (pairs of) fastq files. For Ion Torrent input, unmapped bam files are also allowed. If you have the Pro Linux version of Target HLA, then you can also use sff files as well.

If using paired reads, the two input files are assumed to have the exact matching reads, in the exact same order (there is currently no in-built check for this).

You are also allowed to process **multiple (pairs of) files for a single sample**, if you have a very large dataset for a single sample. Note you can either process multiple files for one sample or multiple samples, but not both at the same time.

You can chose between copying the selected files into Omixon Target's database (by checking the box next to 'Import data into tool') or leaving the selected data in its current disk location (this is the default setting).

A background task will be started, and once it's finished a new item will appear in the main list within the HLA Typing Dashboard where you will be able to see the result.

HLA Typing uses 8-digit typing precision. Depending on the width of coverage of your data (and whether or not candidate alleles have intronic data in the database), results can have 6 or 8 digit resolution.

Protocol

Protocols are stored sets of HLA Typing analysis parameters. Using an existing protocol can speed up the HLA Typing wizard, as most of the parameters will be set to those stored in the protocol.

If you settle on a protocol that works well for your samples, you can save this protocol from any successful run and reuse it again later. You can also edit existing protocols, and create new ones at any time.

Paired File Handling

Automatic pair handling

Paired files can now all be selected together in the 'Select input files' page. Ensure that the 'Paired read import' and 'Automatic pair selection' checkboxes are selected in the 'Paired read options' page.

If there is a problem with the automatic pair selection, you will have to manually tell the tool which is the first paired file and which is the second. In this case, select the first file for each sample in the 'Select input files' page (you can still select multiple files for multiple samples), and then deselect the 'Automatic pair selection' checkbox and choose the second file for each sample in the 'Paired read options' page.

If using paired reads, the two input files are assumed to have the exact matching reads, in the exact same order (there is currently no in-built check for this).

Paired read options

For most Illumina data sets the defaults will be fine.

- Minimum Distance - the minimum expected distance between the two reads
- Maximum Distance - the maximum expected distance between the two reads
- Orientation - the orientation of the reads with respect to each other. Options are FR (forward-reverse), RF (reverse-forward) and FF (forward-forward).

Advanced Options

These options dictate where the data has come from and effect how the underlying algorithms deal with the data. Some data sources (e.g. whole genome and whole exome) are more 'noisy', and the algorithms can help to filter out this extra noise.

- Process all reads. For whole exome or whole genome data sets it is recommended to process all the reads (or pairs) because of the lower coverage that is usually found in these kind of data sets.
- Read processing option. You can now choose from a pre-configured set of typical sequencing runs, including Whole Genome and Whole Exome. You can also choose Custom to manually set how many reads to process. If you are not sure, then use the 'All' option to process all the reads.
- Maximum Reads Processed. **This is very important. You may need to experiment with this value to find the best match for your data.** This dictates how many reads (or how many pairs) will be processed from each input file in order to do the HLA Typing. For very targeted data sets including only the HLA loci, this value doesn't usually need to be too high (about 20'000). For larger kits such as the RainDance HLAseq kit (the whole HLA superregion) this will need to be significantly higher (at least 2'000'000).
- Ignore Cross Mappings. Some sources of data can be more 'noisy', where there are reads from multiple loci that map to each other. Generally, the wider the sequencing, the noisier the data. This option allows reads that have been mapped to multiple HLA loci within the database to be ignored in the results, and acts as a kind of noise filter.
- Ignore Rare Alleles. There is now an option to ignore very rare alleles in the analysis - they will not be included or reported in this case. If this option is not chosen, then these rare alleles will be marked in the results. If this option is greyed out, you will need to 'Setup HLA Typing' again, in order to import the rare alleles database.
- Typing precision. When 8 digit typing is set and both the sequencing and the database have the intronic information, then Omixon Target tries to give an 8 digit result. Otherwise if falls back to 6 digit.
- Save read mappings (bam format). Use this setting to save the mapped reads into a bam format if you would like to browse them in the HLA Genome Browser.

HLA Typing Analysis Result

Introduction

High-level overview of the HLA Typing results

A high-level overview table of the HLA Typing results for one or more submissions of the HLA Typing tool. This can include the summary results for many samples. You can use Ctrl + mouse wheel to zoom in and out of the table.

Short reads are aligned to closely matching allele sequences, then coverage is compared between allele candidates to get the best match(es). The algorithm will automatically select two or more 'best matching' alleles for each locus. These can be manually overridden - 'assigned' - by the user.

Each Sample can be selected, and the following options are available:

- **Paired Results**, all possible pairs of candidates for each locus, select pairs together
- **Sample Details**, the detailed results for a sample
- **Allele Details**, the detailed results for an allele
- **Browse Sample**, a short read alignment visualisation

The results can be exported in TXT (tab delimited text), CSV (comma separated text) or XLS (Excel) format.

Tip - If you would like to export only the assigned allele candidates first click the **Assigned Only** button, then export the results.

At this level, the summary HLA Typing results can be displayed in 4-, 6- or 8-digit formats. In the more detailed results the 6- or 8-digit format can be seen.

Interpreting Results

By default, only the 'best matching' candidates are displayed in the results table. These are the pairs of candidates with best overall combination of 'detection' and 'average coverage' scores. If an ambiguous best match is returned, then there will be more than two best matching candidates for the locus.

By pushing the 'Best Matches Only' button again, it is possible to see other close allele candidates. The 'Setup Filters' option can be used to display more (or less) allele candidates.

By selecting a sample (clicking on the table) you will be presented with more options for drilling down into the details of the results. You can see details for a sample, for a particular allele, you can browse allele candidates in the Genome Browser, and you can see the pair of candidates for each locus. More help is available within each of these functions.

Detection

The proportion of the allele that has sufficient coverage for typing (percentage value). Very uneven coverage will reduce the detection value. Low detection scores will get a yellow warning flag. Very low detection scores will get a red error flag.

Average coverage

The average coverage across the whole allele. High scores are usually better, unless the coverage is very uneven (i.e. lower detection score).

Warning/error flags

Flags might indicate an issue with the results. See the 'Warning Flags' page for more.

Filtering Alleles

A summary of the currently active filters is displayed to the top right of the HLA results screens.

Locus filtering

You can filter the displayed loci by using the **Setup Loci** function.

Status filtering

By clicking the 'Assigned Only' button, unassigned allele candidates can be filtered out. By clicking the 'Best Matches Only' button, lower ranked allele candidates can be filtered out.

Coverage filtering

Using the **Setup Filters** function you can change the filtering options for the main candidates. A minimum average coverage threshold can be set for the summary result table. This limit can be set as an absolute coverage value (i.e. the minimum coverage for the allele) and/or a relative coverage threshold (minimum coverage percentage relative to the 'best matching' allele). The default value for minimum coverage is 10 for the absolute and 95% for the relative coverage filter function. Alleles below the absolute coverage limit are shown as empty cells. Note, that allele candidates with a lower coverage than the selected limit are NOT shown on the "HLA Typing sample result" dashboard, which contains the detailed results.

Imbalanced result filtering

Using the **Setup Filters** function you can change the filtering options for the imbalanced allele candidates. They are marked when a homozygous result has been assigned and we detect at least one more high quality candidate with a very low proportion of the read data compared to the homozygous allele (i.e. imbalanced result). Imbalanced results like these may be due to issues with primers. Using the **Imbalanced Filter** button on the main display these imbalanced candidates can be viewed. The minimum imbalanced coverage filter works exactly like the minimum coverage threshold above, but only applies to the imbalanced allele candidates. The relative imbalanced detection is relative to the best detection score of the 'main' candidates. The minimum imbalanced matches controls how many imbalanced allele candidates we display in the table. Setting this higher will display more candidates.

Noise detection filtering

Although the primary purpose of the imbalanced allele detection is for the detection of primer issues leading to homozygous results (see above), the imbalanced candidates can also be viewed for heterozygous results (press the **Imbalanced Filter** again so that all imbalanced candidates are displayed, not just the homozygous ones). This can be considered a 'noise detection' facility and will not normally be needed. It may be useful in detecting contamination across multiple samples or helping to diagnose very noisy data.

Assigning Alleles

For each allele, one or more allele candidates can be assigned. There are four different ways to assign allele candidates:

- Candidates can be assigned manually by clicking on the check mark before the allele candidate's name.
- All 'best matching' candidates (i.e. the best allele candidates based on coverage and detection statistics) can be assigned using the **Assign Best Matches** button.
- All unambiguous results can be assigned using the **Assign Unambiguous** button.
- All allele candidates can be assigned using the **Assign All** button.

All assignments can be deleted by the **Unassign All** button. Note, that all the 'Assign ...' buttons on this screen affect all the displayed samples and alleles - they don't affect alleles that are not being displayed due to filtering criteria. If you prefer to do the assignments separately for each sample/allele, you can find similar functions by using the **Sample Details** and **Allele Details** buttons.

Warning Flags

- Alleles displayed with the blue font are homozygous.
- Alleles displayed in italics are imbalanced (i.e. present in a much smaller proportion of the reads that the main candidates).

Results are annotated with warning flags. Some of these flags apply to single allele candidates, others apply to the pair of candidates at a particular locus.

Single candidate flags:

- Allele candidates with lower than 90% 'detection' on exons 2 and 3 (Class I genes) or exon 2 (Class II genes) get an 'error flag' (red). The 'detection' is the proportion of the exon with sufficient coverage. This flag indicates that there may be some key areas of the candidate not (well) covered by short read data, so the result should be treated with caution and may require manual inspection.
- Allele candidates with lower than 98% 'detection' on exons 2 and 3 (Class I genes) or exon 2 (Class II genes) get a 'warning flag' (orange).
- If rare alleles are not ignored (option during the analysis), then they will be marked with a red exclamation mark in the results.
- A yellow 'balance' icon next to a homozygous result indicates a possible imbalanced allele in the result. This is a warning that there may be an issue with the primer, and one of the two candidate alleles is present in a very low proportion of the read data compared to the other (i.e. imbalanced result). Using the imbalanced filter these imbalanced candidates can be viewed.

Paired candidate flags:

- Where there are multiple allele candidate pairs with very similar coverage and detection statistics, these will be marked with a purple pair flag. This indicates that - while these candidates are not 100% ambiguous (i.e. identical scores) - they are so close that it is difficult to tell them apart, and manual inspection of the results may be required.
- Pairs marked with a red pair flag have too many candidates in the list of possible pairs. This maybe an indication of a problem in the sequencing, either with contamination or demultiplexing.
- Pairs marked with a yellow pair flag have a few too many candidates in the list of possible pairs. This is a warning flag.

HLA Typing Sample Result

A more detailed view of the HLA Typing results for a single Sample

Short reads are aligned to closely matching allele sequences, then coverage is compared between allele candidates to get the best match(es).

For each allele at each locus we supply:

- Detection (%) - Proportion of the allele with sufficient coverage. The allele uses the lowest detection value among its exons. High values are better.
- Average coverage - The average number of short reads covering the whole allele (exons with zero coverage are not counted).
- Exons covered - Number of exons with non-zero coverage / number of available exon sequences for the allele candidate.

By clicking on the Browse button next to the name of a specific allele, the **HLA genome browser** can be opened. This HLA visualisation tool shows the short reads in the Sample aligned to the allele candidates. The best allele candidate is marked as "Best Match".

By clicking on the Analyse button next to the allele name, the **HLA Typing allele result** screen is opened. This page contains detailed, exon level statistics for the allele candidates.

A minimum coverage threshold can be set for the summary result table. This limit can be set as an absolute coverage value (i.e. the mean number of reads for the allele) or a relative coverage threshold (minimum coverage relative to the top allele). The default value for minimum coverage is 10 for the absolute and 95% for the relative coverage filter function. Alleles below the coverage limit are shown as empty cells. Note, that allele candidates with a lower coverage than the selected limit are NOT shown on the "HLA Typing sample result" dashboard, which contains the detailed results.

For each allele, one or more allele candidates can be assigned. There are four different ways to assign allele candidates:

1. Candidates can be assigned manually by clicking on the checkmark before the allele candidate's name.
2. All 'best matching' candidates (i.e. the best allele candidates based on coverage statistics) can be assigned using the **Assign Best Matches** button.
3. All unambiguous results can be assigned using the **Assign Unambiguous** button.
4. All allele candidates can be assigned using the **Assign All** button.

All assignments can be deleted by the **Unassign All** button. Note, that the 'Assign ...' buttons on this screen affect all the displayed samples, if you prefer to do the assignments separately for each sample/allele, you can find similar functions on the HLA Typing sample and allele result screens.

The results can be exported in TXT (tab delimited text), CSV (comma separated text) or XLS (Excel) format.

Tip - If you would like to export only the assigned allele candidates click the **Assigned Only** button, then export the results.

You can filter the displayed loci by using the **Setup Loci** function.

Warning Flags

- Alleles displayed with the blue font are homozygous.
- Alleles displayed in italics are imbalanced (i.e. present in a much smaller proportion of the reads that the main candidates).

Results are annotated with warning flags. Some of these flags apply to single allele candidates, others apply to the pair of candidates at a

particular locus.

Single candidate flags:

- Allele candidates with lower than 90% 'detection' on exons 2 and 3 (Class I genes) or exon 2 (Class II genes) get an 'error flag' (red). The 'detection' is the proportion of the exon with sufficient coverage. This flag indicates that there may be some key areas of the candidate not (well) covered by short read data, so the result should be treated with caution and may require manual inspection.
- Allele candidates with lower than 98% 'detection' on exons 2 and 3 (Class I genes) or exon 2 (Class II genes) get a 'warning flag' (orange).
- If rare alleles are not ignored (option during the analysis), then they will be marked with a red exclamation mark in the results.
- A yellow 'balance' icon next to a homozygous result indicates a possible imbalanced allele in the result. This is a warning that there may be an issue with the primer, and one of the two candidate alleles is present in a very low proportion of the read data compared to the other (i.e. imbalanced result). Using the imbalanced filter these imbalanced candidates can be viewed.

Paired candidate flags:

- Where there are multiple allele candidate pairs with very similar coverage and detection statistics, these will be marked with a purple pair flag. This indicates that - while these candidates are not 100% ambiguous (i.e. identical scores) - they are so close that it is difficult to tell them apart, and manual inspection of the results may be required.
- Pairs marked with a red pair flag have too many candidates in the list of possible pairs. This maybe an indication of a problem in the sequencing, either with contamination or demultiplexing.
- Pairs marked with a yellow pair flag have a few too many candidates in the list of possible pairs. This is a warning flag.

HLA Typing Allele Result

Allele candidate level HLA typing results

A more detailed view of the HLA Typing results for a single Allele candidate.

Short reads are aligned to closely matching allele sequences, then coverage is compared between allele candidates to get the best match(es).

For each exon of each allele candidate we supply:

- Detection (%) - Proportion of the exon with sufficient coverage. High values are better.
- Average coverage - The average number of short reads covering the whole exon or allele (exons with zero coverage are not counted).
- Exons covered - Number of exons with non-zero coverage / number of available exon sequences for the allele candidate.

A minimum coverage threshold can be set for the allele result table. This limit can be set as an absolute coverage value (i.e. the mean number of reads for the allele) or a relative coverage threshold (minimum coverage relative to the top allele). The default value for minimum coverage is 10 for the absolute and 95% for the relative coverage filter function. Alleles below the coverage limit are shown as empty cells. Note, that allele candidates with a lower coverage than the selected limit are NOT shown on the "HLA Typing allele result" dashboard, which contains the detailed results.

For each allele, one or more allele candidates can be assigned. There are four different ways to assign allele candidates:

1. Candidates can be assigned manually by clicking on the checkmark before the allele candidate's name.
2. All 'Best matching' candidates (i.e. the best allele candidates based on coverage statistics) can be assigned using the **Assign Best Matches** button.
3. All unambiguous results can be assigned using the **Assign Unambiguous** button.
4. All allele candidates can be assigned using the **Assign All** button.

All assignments can be deleted by the **Unassign All** button.

The results can be exported in TXT (tab delimited text), CSV (comma separated text) or XLS (Excel) format.

Tip - If you would like to export only the assigned allele candidates click the **Assigned Only** button, then export the results.

By clicking on the **Browse results** button, the **HLA genome browser** can be opened. This HLA visualisation tool shows the short reads in the Sample aligned to the allele candidates. The best allele candidate is marked as 'Best Match'.

Warning Flags

- Alleles displayed with the blue font are homozygous.
- Alleles displayed in italics are imbalanced (i.e. present in a much smaller proportion of the reads that the main candidates).

Results are annotated with warning flags. Some of these flags apply to single allele candidates, others apply to the pair of candidates at a particular locus.

Single candidate flags:

- Allele candidates with lower than 90% 'detection' on exons 2 and 3 (Class I genes) or exon 2 (Class II genes) get an 'error flag' (red). The 'detection' is the proportion of the exon with sufficient coverage. This flag indicates that there may be some key areas of the candidate not (well) covered by short read data, so the result should be treated with caution and may require manual inspection.
- Allele candidates with lower than 98% 'detection' on exons 2 and 3 (Class I genes) or exon 2 (Class II genes) get a 'warning flag' (orange).
- If rare alleles are not ignored (option during the analysis), then they will be marked with a red exclamation mark in the results.
- A yellow 'balance' icon next to a homozygous result indicates a possible imbalanced allele in the result. This is a warning that there may be an issue with the primer, and one of the two candidate alleles is present in a very low proportion of the read data compared to the other (i.e. imbalanced result). Using the imbalanced filter these imbalanced candidates can be viewed.

Paired candidate flags:

- Where there are multiple allele candidate pairs with very similar coverage and detection statistics, these will be marked with a purple pair flag. This indicates that - while these candidates are not 100% ambiguous (i.e. identical scores) - they are so close that it is difficult to tell them apart, and manual inspection of the results may be required.
- Pairs marked with a red pair flag have too many candidates in the list of possible pairs. This maybe an indication of a problem in the sequencing, either with contamination or demultiplexing.
- Pairs marked with a yellow pair flag have a few too many candidates in the list of possible pairs. This is a warning flag.

Paired Result

In this screen you can view and assign HLA typing results by seeing the **most likely pairs of alleles per locus**.

This screen shows the top pairs of allele candidates, and allows one or more pairs to be selected together.

Results are annotated with warning flags:

- Loci with at least two, similarly 'strong' candidate pairs get an 'ambiguity flag' (purple). This indicates that - while these candidates are not 100% ambiguous (i.e. with completely identical scores) - they are so close that it is difficult to tell them apart, and manual inspection of the results may be required.
- Loci with more than 50 candidate allele pairs get an 'error flag' (red). This may be an indication of a problem in the sample prep or sequencing, either with contamination or demultiplexing.
- Loci with more than 30 candidate allele pairs get a 'warning flag' (yellow). This is a warning flag.

Other indicators:

- Alleles displayed with a blue font are homozygous.
- Alleles displayed in italics are imbalanced (i.e. present in a much smaller proportion of the reads that the main candidates).

HLA Genome Browser

Introduction

The HLA Genome Browser allows visual inspection of genomics data.

Multiple allele candidates can be browsed together.

The data items that can be visualised for each allele candidate include:

- Reference Sequence
- Exon Annotations
- Short Reads
- Coverage

By default, short read alignments are displayed in a stranded fashion, so that forward strand reads (pink) and reverse strand reads (yellow) can be easily distinguished within the display. The 'Display Strand' function turns this on and off.

The short reads track can be 'collapsed' which gives a summary view of the short reads (and does not allow each read to be inspected in detail).

Individual items (exons, short reads) within the HLA Genome Browser can be selected, and it's possible to browse elsewhere within the same locus and then jump back to the selected item, or copy the details of the selected item to the clipboard.

Settings and Functions

Expanding the display

At the top left of the tool is a small control that allows an expanded view. This actually works for all screens in the tool, but is most useful for the HLA Genome Browser. Clicking the same icon again will go back to the usual display.

Rotating the Genome Browser

The vertical HLA Genome Browser view is more useful for comparing multiple allele candidates. The view can be rotated to use a horizontal display, which is a bit more traditional for Genome Browsers, and is more useful when browsing a single allele candidate.

Zooming

By default the Genome Browser starts in 'Drag' mode, where the mouse wheel (or '+' and '-' keys) can be used to zoom in and out, and the display can be moved by clicking and dragging with the mouse. 'Drill' mode can also be used, where the region highlighted by a mouse click and drag will be 'drilled into' once the mouse button is released.

Jumping Around

You can jump to a position in the reference using 'Jump To Position'. If you have previously selected a feature within the display (such as an annotation or short read), then you can jump back to that feature at any time by using the small arrow button at the bottom of the display.

Search for a nucleotide sequence

By pressing CTRL+S, you can open the 'Search sequence' wizard. Here you can specify a nucleotide sequence to search for in the currently visible allele sequences and then step through the hit positions.

Filtering the Short Reads

You can display more short reads by using the 'collapsed' pile-up view.

You can also 'page' around the short reads using the small arrows at the top of the short read track. You can set the page size using the up and down arrows, and the current page by using the left and right arrows. This is useful when you have deep sequencing and want to scroll through a few hundred short reads at a time.

Filtering the Allele Candidates

The list of allele candidates shown in the HLA Genome browser can be filtered by coverage. This limit can be set as an absolute coverage value (i.e. the mean number of reads for the allele) or a relative coverage threshold (minimum coverage relative to the top allele). The default value for minimum coverage is 10 for the absolute and 95% for the relative coverage filter function. Current coverage filter settings can be seen in the top right corner, above the alignments.

By clicking the 'Assigned Only' button, unassigned allele candidates can be hidden. By clicking the 'Best Matches Only' button, lower ranked allele candidates can be hidden.

Analysing result

By clicking on the 'Analyse Result' button, the HLA Typing allele result screen is opened. This page contains detailed, exon level statistics for the allele candidates. Note, that only samples visible in the current browser session will be displayed in the allele result table.

Track Setup

You can configure which tracks you would like to see in the display.

Display Setup

You can display/hide read pair information, indels, SNPs and soft-clips.

Exporting HLA Genome Browser Data

You can 'Copy to Clipboard' the details of individual, selected items in the display. You can also copy the reference sequence belonging to a selected item (read or annotation), by clicking on the 'Copy reference to clipboard' button.

You can also 'Capture a Screenshot' of the currently visibly tracks within the HLA Genome Browser. This will create a .png file at your chosen location.

Context menu

By clicking the right mouse button, you can open the Context menu. In the HLA Browser, this menu contains the following functions:

- Search sequence: with this function you can search for a specific nucleotide sequence in the reference of all visible alleles.
- Toggle fullscreen: you can toggle between full screen (i.e. only the browser is shown) and the normal view.
- Toggle reference masked: with this function, you can hide all the bases shared by the different allele reference sequences on the screen. This way, only the differences between the allele candidates are shown.
- Add custom candidate: you can add any alleles of the currently selected gene from hla.dat. Reads that can be aligned to this custom allele will be shown in the browser. Note, that allele candidates that are already in the candidate list for that gene cannot be added again.
- Remove custom candidate: the previously added custom allele can be removed by this function.

Shortcuts

Keyboard/mouse navigation options and shortcuts

- **UP/DOWN arrow keys**: scroll alongside the reference contig.
- **LEFT/RIGHT arrow keys**: scroll through the reads at the current position/area.
- **PAGE UP/PAGE DOWN**: scroll alongside the reference contig in bigger steps.
- **HOME/END**: jump to the beginning/end of the current reference contig.
- **+/-**: zoom in and out.
- **SPACE**: locks/unlocks the cursor (i.e. green line when unlocked, red line when locked).
- **BACKSPACE**: shows the whole reference contig (i.e. zooms out totally) or when the cursor is locked, jumps to the locked cursor.
- **Mousewheel**: zoom in/out (in 'Drag' mode) or move alongside the reference (in 'Drill' mode).\
- **SHIFT+mousewheel**: scroll through the reads at the current position/area.
- **Left mouse button**: can be used for dragging the display (in 'Drag' mode) or for highlighting and zooming in on a region (in 'Drill' mode by clicking and releasing the left button).
- **CTRL+J**: opens the 'Jump to position' wizard. Here you can specify a reference contig and a coordinate or area on that contig and jump to the selected location.
- **CTRL+S** : opens the 'Search sequence' wizard. Here you can specify a nucleotide sequence to search for in the currently visible reference contig and then step through the hit positions.
- **CTRL+M**: enter full screen mode - hide all menus and show the browser region in full screen.
- **ESCAPE**: exit full screen mode.
- **X**: switches the browser between horizontal and vertical mode.
- **C**: switches the browser between collapsed and uncollapsed mode.
- **T**: switches the browser between drag and drill navigation mode.

HLA Manage Tracks

In this wizard, you can hide and show any tracks.

The visibility of any track can be set with the 'Show Track(s)' and 'Hide Track(s)' functions. Hidden tracks are marked with a darker grey color in the 'Manage Tracks' wizard track list.

HLA Export Results

HLA typing results can be exported by clicking on the "Export Results" button on the HLA Typing dashboard. This function generates separate xml formatted files for each sample in the selected Analysis or Analyses. These results can be later imported back into Omixon Target, using the "Import HLA typing results" function on the left panel.

HLA Typing Import

Finished HLA typing results can be exported in xml format. These result files can later be imported back into Omixon Target, using the "Import HLA typing results" function on the left panel. To import the results, just check a list of result files and a new Analysis will automatically be created for the imported results. Note, that these files don't contain alignment information, so you won't be able to use the HLA Genome Browser for these samples.

HLA Export Table

Every HLA typing result table can be exported from Omixon Target in either csv, txt (tab delimited text), xls or xlsx format. Note, that only visible results are exported (i.e. if an allele is filtered out based on low coverage, it won't be shown in the exported file either).

There are several similar export table functions in the HLA module of Omixon Target:

- On the 'HLA Typing dashboard', you can export an overview of all analyses. (Note, that the "Table Export" function is only available when the Display mode is set to table.)
- On the 'HLA Typing analysis results' dashboard, you can export all the results from an analysis. If you select the "Overview" export, only the allele names will be exported. The "Detailed" table also contains flags and statistics for each allele candidate.
- On the 'HLA Typing sample result' page, you can export data for a single allele (by pressing the small Export button in the allele header).
- On the 'HLA Typing allele result' page, you can export detailed statistics of the allele candidates of a single allele.

HLA Typing Protocol Dashboard

Introduction

This function is available from the main HLA Typing Dashboard. You can view, edit and use protocols from here.

Protocols are stored sets of HLA Typing analysis parameters. Using an existing protocol can speed up the HLA Typing wizard, as most of the parameters will be set to those stored in the protocol.

If you settle on a protocol that works well for your samples, you can save this protocol from any successful run and reuse it again later. You can also edit existing protocols, and create new ones at any time.

The default protocol cannot be deleted or edited, it can be copied.

Clicking on 'use' for any protocol will launch the main HLA Typing wizard using the parameters from within the selected protocol.

HLA Typing Protocol

Introduction

This function is available from the HLA Typing Protocols Dashboard. You can view, edit and use protocols from here.

Protocols are stored sets of HLA Typing analysis parameters. Using an existing protocol can speed up the HLA Typing wizard, as most of the parameters will be set to those stored in the protocol.

If you settle on a protocol that works well for your samples, you can save this protocol from any successful run and reuse it again later. You can also edit existing protocols, and create new ones at any time.

Paired Read Options

If using paired reads, the two input files are assumed to have the exact matching reads, in the exact same order (there is currently no in-built check for this).

Paired read options

For most Illumina data sets the defaults will be fine.

- Minimum Distance - the minimum expected distance between the two reads
- Maximum Distance - the maximum expected distance between the two reads
- Orientation - the orientation of the reads with respect to each other. Options are FR (forward-reverse), RF (reverse-forward) and FF (forward-forward).

Advanced Options

These options dictate where the data has come from and effect how the underlying algorithms deal with the data. Some data sources (e.g. whole genome and whole exome) are more 'noisy', and the algorithms can help to filter out this extra noise.

- Process all reads. For whole exome or whole genome data sets it is recommended to process all the reads (or pairs) because of the lower coverage that is usually found in these kind of data sets.
- Read processing option. You can now choose from a pre-configured set of typical sequencing runs, including Whole Genome and Whole Exome. You can also choose Custom to manually set how many reads to process. If you are not sure, then use the 'All' option to process all the reads.
- Maximum Reads Processed. **This is very important. You may need to experiment with this value to find the best match for your data.** This dictates how many reads (or how many pairs) will be processed from each input file in order to do the HLA Typing. For very targeted data sets including only the HLA loci, this value doesn't usually need to be too high (about 20'000). For larger kits such as the RainDance HLAseq kit (the whole HLA superregion) this will need to be significantly higher (at least 2'000'000).
- Ignore Cross Mappings. Some sources of data can be more 'noisy', where there are reads from multiple loci that map to each other. Generally, the wider the sequencing, the noisier the data. This option allows reads that have been mapped to multiple HLA loci within the database to be ignored in the results, and acts as a kind of noise filter.
- Ignore Rare Alleles. There is now an option to ignore very rare alleles in the analysis - they will not be included or reported in this case. If this option is not chosen, then these rare alleles will be marked in the results. If this option is greyed out, you will need to 'Setup HLA Typing' again, in order to import the rare alleles database.
- Typing precision. When 8 digit typing is set and both the sequencing and the database have the intronic information, then Omixon Target tries to give an 8 digit result. Otherwise if falls back to 6 digit.
- Save read mappings (bam format). Use this setting to save the mapped reads into a bam format if you would like to browse them in the HLA Genome Browser.

Sandbox Browser

Sandbox Genome Browser

Introduction

The Sandbox Genome Browser function is for quick visualisation of alignments, variants and annotations. This module of Omixon Target is available free-of-charge and can be used as a standalone genome browser.

The Sandbox Genome Browser was designed with "ease of use" in mind. Multiple alignments, variant calls and annotation files can be imported within a single wizard. Multiple browsing sessions (i.e. basically sets of tracks with visual settings) can be saved. Saved sessions can be reopened with a few clicks and can be edited freely at any time. The last used browsing session is stored and reopened

automatically when Omixon Target is restarted.

The browser comes with a built-in HG19 reference, so if the selected alignment, annotation or variant call files are HG19 based, the required reference chromosomes are automatically detected and loaded.

Settings and Functions

Expanding the display

At the top left of the tool is a small control that allows an expanded view. This actually works for all screens in the tool, but is most useful for the Genome Browser. Clicking the same icon again will go back to the usual display.

Rotating the Genome Browser

The vertical Genome Browser view is more useful for comparing multiple samples. The view can be rotated to use a horizontal display, which is a bit more traditional for Genome Browsers, and is more useful when browsing a single sample.

Filtering the Short Reads

You can display more short reads by using the 'collapsed' pile-up view.

In collapsed mode, you can also 'page' around the short reads using the small arrows at the top of the short read track. You can set the page size using the up and down arrows, and the current page by using the left and right arrows. This is useful when you have deep sequencing and want to scroll through a few hundred short reads at a time.

Display setup

The visualisation of variants in the short read track can be customized as well. Variants are always visible on the nucleotide level, but insert and deletion marks can be selectively turned on and off. Soft clips can be shown or hidden as well. Reads can be shown in normal mode or paired mode (in paired mode, reads are sorted differently and pairs are connected with a thin line).

Exporting Meta-data and Creating Screenshots

You can 'Copy to Clipboard' the details of individual, selected items in the display. You can select a read and by clicking on the 'Copy to clipboard' button (which can be found on the middle panel in the bottom of your screen), you can copy the name and CIGAR string of the selected read to the clipboard. Similarly, for a selected annotation, the position and name of the annotation can be copied to the clipboard.

You can also copy the reference sequence belonging to a selected item (read or annotation), by clicking on the 'Copy reference to clipboard' button.

You can also 'Capture a Screenshot' of the currently visibly tracks within the Genome Browser. This will create a .png file at your chosen location.

Working with Tracks

Tracks used in the Sandbox Genome Browser

There are two different kinds of tracks used within the Sandbox Genome Browser. **"Built-in tracks"** can be opened via the 'Open...' function found on the side panel. These tracks are not editable and are automatically added to the Omixon Target database. The other kind of tracks are called **"User tracks"**. These tracks can be editable and are stored in external (non database) files.

Creating tracks

User tracks can be created within Omixon Target (using the 'Create Track' button in the 'Manage Tracks' wizard), or can be imported from external files (using the 'Add Track' function in the 'Manage Tracks' wizard). User tracks created within Omixon Target are always editable. User tracks imported from file can be editable or non-editable (you can set this when importing the file). Annotation editing is currently only supported for files in bed format.

Modifying tracks

Currently, track editing is only available for bed formatted, editable, user defined tracks.

- ***To create a new annotation entry:*** use the green '+' button on the middle part of the bottom panel, or simply press the 'Insert' button on your keyboard. This will automatically generate an annotation entry based on the currently shown browser region. You can choose which track you want to add this newly defined entry to, from a list containing all the editable tracks and you can also modify the name and end coordinates of the annotation entry by hand.
- ***To modify an existing annotation entry "by hand":*** you have to select it first (this can be done simply by clicking on the entry with the left mouse button). The selected annotation element is marked by a lighter colour. The attributes of the annotation element can be edited using the 'Edit selected annotation' function (this button is located on the middle bottom panel, next to the 'Create new annotation' button). You can also open the annotation editor screen by pressing the ENTER button, after selecting an annotation.
- ***To resize an annotation entry in the browser*:** select the annotation, lock the cursor (you can do this by pressing 'Space' on

your keyboard), then press the 'Resize selected annotation' button on the middle bottom panel or press 'R' on the keyboard. The selected annotation will be resized, to end at the cursor position.
- ***To move an annotation entry in the browser:*** select the annotation, lock the cursor (you can do this by pressing 'Space' on your keyboard), the press the 'Move selected annotation' button on the middle bottom panel or press 'M' on the keyboard. The selected annotation will be moved to the cursor position.

Deleting tracks

To delete an annotation entry, select the annotation, then press the 'Delete selected annotation' button (red 'X' symbol on the middle bottom panel) or the 'Delete' button on your keyboard.

Showing or hiding tracks

All tracks (both built-in and user created tracks) can be temporarily hidden using the 'Hide track(s)' button in the 'Track Manager' wizard.

Navigation in the Browser

Zooming

By default the Genome Browser starts in 'Drag' mode, where the mouse wheel (or '+' and '-' keys) can be used to zoom in and out, and the display can be moved by clicking and dragging with the mouse. 'Drill' mode can also be used, where the region highlighted by a mouse click and drag will be 'drilled into' once the mouse button is released.

Jump to a position

You can jump to a position in the reference using 'Jump To Position', or jump to one of the annotations within the Target using 'Jump To Target' or 'Jump to annotation'. If you have previously selected a feature within the display (such as an annotation or short read), then you can jump back to that feature at any time by using the small arrow button at the bottom of the display.

The 'Jump To Position' wizard can be opened by pressing the "CTRL" and "J" keys on the keyboard. After the wizard is opened, you can type in a single coordinate or a pair of coordinates and you can specify a reference contig. When a reference contig is not chosen, the currently shown contig is used. Coordinates can be specified with or without thousand delimiters (i.e. 15200, 15.200 and 15,200 are all accepted and mean the same).

Some examples (the currently selected reference contig is chr1):

- ***CTRL+J 15200 ENTER*** -> The Genome Browser display will jump to chr1:15200.
- ***CTRL+J 15.200-15.400 ENTER*** -> The Genome Browser display will show the area chr1:15200 - 15400 wizard can be opened by pressing the "CTRL" and "J" keys on the keyboard. After the wizard is opened, you can type in a single coordinate or a pair of coordinates and you can specify a reference contig. When a reference contig is not chosen, the currently shown contig is used. Coordinates can be specified with or without thousand delimiters (i.e. 15200, 15.200 and 15,200 are all accepted and mean the same thing).
- ***CTRL+J chr2:15,200 ENTER*** -> The Genome Browser display will jump to chr2:15200.

Jump to an annotation

The 'Jump To Annotation' wizard can be opened by pressing the "CTRL" and "F" keys on the keyboard. After the wizard is opened, you can type in the name (or the beginning of the name) of an annotation (e.g. gene, exon, amplicon). When the specified search term is ambiguous (i.e. more than one annotation name starts with the specified string), a list of annotations is presented and the desired annotation can be selected from the list. The input field in this wizard is not case sensitive (i.e. brca1 and BRCA1 are both accepted and mean the same thing).

Search for a nucleotide sequence

By pressing CTRL+S, you can open the 'Search sequence' wizard. Here you can specify a nucleotide sequence to search for in the currently visible reference contig and then step through the hit positions.

Shortcuts

Keyboard/mouse navigation options and shortcuts

- **UP/DOWN arrow keys**: scroll alongside the reference contig.
- **LEFT/RIGHT arrow keys**: scroll through the reads at the current position/area.
- **PAGE UP/PAGE DOWN**: scroll alongside the reference contig in bigger steps.
- **HOME/END**: jump to the beginning/end of the current reference contig.
- **+/-**: zoom in and out.
- **SPACE**: locks/unlocks the cursor (i.e. green line when unlocked, red line when locked).
- **BACKSPACE**: shows the whole reference contig (i.e. zooms out totally) or when the cursor is locked, jumps to the locked cursor.
- **Mousewheel**: zoom in/out (in 'Drag' mode) or move alongside the reference (in 'Drill' mode).
- **SHIFT+mousewheel**: scroll through the reads at the current position/area.
- **Left mouse button**: can be used for dragging the display (in 'Drag' mode) or for highlighting and zooming in on a region (in 'Drill' mode by clicking and releasing the left button).
- **TAB/SHIFT+TAB**: jump to next/previous reference contig.

- **CTRL+J**: opens the 'Jump to position' wizard. Here you can specify a reference contig and a coordinate or area on that contig and jump to the selected location.
- **CTRL+F**: opens the 'Jump to annotation' wizard. Here you can specify the name (or partial name) of an annotation (e.g. gene, amplicon, exon) and jump to the selected annotation.
- **CTRL+S** : opens the 'Search sequence' wizard. Here you can specify a nucleotide sequence to search for in the currently visible reference contig and then step through the hit positions.
- **CTRL+M**: enter full screen mode - hide all menus and show the browser region in full screen.
- **ESCAPE**: exit full screen mode.
- **X**: switches the browser between horizontal and vertical mode.
- **C**: switches the browser between collapsed and uncollapsed mode.
- **T**: switches the browser between drag and drill navigation mode.
- **INSERT**: insert new annotation.
- **ENTER**: edit selected annotation.
- **DELETE**: delete selected annotation.
- **M**: move selected annotation to start at the locked cursor position.
- **R**: resize selected annotation to end at the locked cursor position.

Sandbox Open File

In this wizard, you can visualise multiple alignments and create an annotation and variant track in a single step.

Reference files must be in fasta (or multifasta) format. Note, that a built-in HG19 reference is available within the Sandbox Genome Browser, so if your alignments, annotations or variant calls are HG19 based, no reference file is need to be imported, you simply have to select a list of alignment and/or annotation and/or variant call files and open them all in one step.

Annotations must be in gff or bed format. Annotation files selected in this wizard are merged and visualised in a single track. (For visualising multiple tracks separately, use the 'Track Manager' function).

Alignment files must be in sam or bam format, unsorted and/or unindexed files are also accepted and are automatically sorted and/or indexed during the import.

Tracks imported in this wizard are called **"built-in tracks"**, these tracks are automatically imported into the Omixon Target database and are not editable.

Sandbox Manage Files

On this screen, you can manage files selected in the 'Open...' wizard. Both reference files (reference contigs and annotations) and sample files (alignments and variant calls) can be deleted individually. The order of sample files can be changed as well, by selecting a single alignment or variant call and moving it up or down in the list.

You can also create new sessions, using the 'Save as...' function. This will basically create a duplicate of the current session (including the current browsing location) with a new name. This function can be used for saving interesting or problematic regions for further analysis or simply for archival reasons.

Tip: For temporarily hiding tracks, see the 'Manage Tracks' function.

Sandbox Manage Sessions

In this wizard, you can create new (empty) sessions and delete, copy, rename or open existing sessions.
The current session is not deletable and at least one session should exist at all times (this means, that if you have a single session, this session can not be deleted).
Tip: The current session is marked with a yellow frame.

Sandbox Manage Tracks

In this wizard, you can hide and show any tracks, and import, create and delete additional annotation and variant tracks.

The visibility of any track can be set with the 'Show Track(s)' and 'Hide Track(s)' functions. Hidden tracks are marked with a darker grey color in the 'Manage Tracks' wizard track list.

Annotation tracks created by hand or added from file in this wizard are called **"User tracks"**. These tracks can be edited within the Genome Browser.

Empty tracks can be created using the 'Create Track' button. These tracks are always editable.

Predefined (i.e. non-empty) tracks can be imported from file using the 'Add Track' button. Annotations must be in gff or bed format, variant calls are accepted in vcf format. The editability of these tracks can be set during the import process. For non-empty tracks, the type of the imported annotations (variant, exon, gene or amplicon) can be set as well. The default annotation type is "exon" for gff and

bed files and "variant" for vcf files.

Sandbox Create Track

In this wizard, an empty annotation track can be created within Omixon Target. The name and type of the annotation can be set (the annotation type can be exon, gene or amplicon). This user defined annotation track is stored in a bed formatted file, at a location selected by the user. When the annotation track is edited within the Genome Browser, this storing bed file is automatically updated and saved.

Sandbox Add Track

In this wizard, annotation tracks (in vcf, gff or bed format) can be imported from external files. The type of the imported track (variant, exon, gene or amplicon) and the 'Fallback track name' (i.e. the name shown in the Browser track header) can be specified. For vcf files, the track type is automatically set to 'Variant annotation'. The editability of the imported track can be set as well. (Note, that editing is only supported for bed formatted files at the moment.)

Settings

Settings Dashboard

Reachable from the Home Dashboard, the Settings Dashboard displays an overview of the settings in the tool, and allows access to the user management features, the visual display settings, license settings and the bug report (mail) settings.

There is also the (dangerous!) 'Reset Everything' button. This will cause the whole Application to be reset - i.e. **this will delete all imported and analysed data**, and reset the whole of Omixon Target back to a 'brand new' status. This function is essentially only recommended for demonstrations, or if you really do want to start again from scratch.

Application Configuration

This wizard allows you to set up a mail server via SMTP, for sending bug and error reports directly to Omixon.

Settings selected in the Application Configuration wizard are saved and will still be in effect after Omixon Target is restarted.

User Management

User Management allows Users to be added, edited and removed from the system.

There must be at least one 'Super User' at all times. The first registered user automatically becomes the Super User, this user cannot be deleted.

Genome Browser Setup

You can define the starting (i.e. used when the software is freshly started) graphical and navigation settings for the Genome Browser (in the Data Analysis module), the free Sandbox Genome Browser and the HLA Browser here. These settings can be saved and will still be in effect after Omixon Target is restarted. The genome browser settings are user specific.

Display configuration

On this screen the orientation of the browser screen (horizontal or vertical), the navigation mode ("Drag & Zoom" or "Drill & Scroll")

Color setup

On this screen, the color codes for nucleotides used in the genome browser can be changed.

Data options

In this part of the wizard you can set whether certain groups of variants should be shown or hidden, based on variant status (accepted, rejected, pending or all) or variant type (insert or deletion). Soft clips can be shown or hidden as well.

Advanced settings

The "zooming resolution" of the Genome Browser can also be set on this screen by changing the 'Overview factor'. This is a the maximum number of nucleotides displayed per pixel. The default setting is 20. The minimum number of pixels for collapsed read width can be set as well.

Tip: For further details about the Display and Data settings, see the Genome Browser, Sandbox Genome Browser or HLA Browser help pages!

Upload license

You can upload a license with this wizard. Just select the file you've got from Omixon and finish the wizard.

The license can be an upgrade from an evaluation license to a full license (either permanent license or annually renewable), or could just be a top-up license for the credit based HLA Typing.

Export Log

This function will export the Omixon Target log file. This is useful if you have encountered a problem within the tool and would like to notify Omixon support staff of the issue - they will usually ask for a log file so that they can investigate your issue and ensure that it is resolved as soon as possible.

Export Audit Log

This function exports the full audit log of everything that has happened to each and every sample inside Omixon Target.

This is useful for audit purposes.

Reset Everything

It is possible to Reset everything - reset the whole of Omixon Target - to an empty state.

Resetting will delete all imported and analysed data.

Resetting is only recommended if it's the first thing that is done with the tool, if the tool is being used for demonstration purposes, or if you really do want to start again completely from scratch.

Reset Everything is available from the Settings Dashboard. You will be asked again if you are completely sure before the reset process is started. There is no way to recover data that is removed by a reset.

Once you have reset, the next recommended step is to use 'Add Profile' to add one (or more) of the pre-configured Profiles to the tool, which will download and import all the data required for each Profile.

Appendix

Keyboard shortcuts

The following generic keyboard shortcuts are available in Omixon Target.

- **F1** key opens the Help for the current page or wizard.
- **F8** key closes the window and exits the application
- **F11** key switches to fullscreen display mode to utilize as much space as possible for visualization - note that certain platforms and window managers are not working properly together with this function, the first time you use you will get a warning message about this.
- **ALT+F4** closes Omixon Target.
- **CTRL + C/V/X** copies/pastes/deletes text in text boxes.
- **SHIFT/CTRL + left click** can be used for multiple selection/deselection.